PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification : Not classified

A2

(11) International Publication Number: **WO 98/38846**

(43) International Publication Date: 11 September 1998 (11.09.98)

(21) International Application Number: PCT/US98/04571

(22) International Filing Date: 6 March 1998 (06.03.98)

(30) Priority Data:

08/813,159	7 March 1997 (07.03.97)	US
60/042,125	28 March 1997 (28.03.97)	US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application

US 60/042,125 (CIP)
Filed on 28 March 1997 (28.03.97)

(71) Applicant *(for all designated States except US)*: AFFYMETRIX, INC. [US/US]; 3380 Central Expressway, Santa Clara, CA 95051 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: LIPSHUTZ, Robert, J. [US/US]; 970 Palo Alto Avenue, Palo Alto, CA 94301 (US). CHEE, Mark [AU/US]; 3199 Waverly Street, Palo Alto, CA 94306 (US). FAN, Jian-Bing [CN/US]; 275 Ventura Avenue #20, Palo Alto, CA 94306 (US). BERNO, Anthony [CA/US]; 570 South 12th Street, San Jose, CA 95112 (US).

(74) Agents: LIEBESCHUETZ, Joe et al.; Townsend and Townsend and Crew LLP, 8th floor, Two Embarcadero Center, San Francisco, CA 94111-3834 (US).

(81) Designated States: JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published

With declaration under Article 17(2)(a).
Without classification and without abstract; title not checked by the International Searching Authority.

(54) Title: GENETIC COMPOSITIONS AND METHODS

AP

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

GENETIC COMPOSITIONS AND METHODS

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation-in-part of USSN 08/813,159, filed March 7, 1997 and USSN 60/042,125 filed March 28, 1997, which are incorporated by reference in their entirety for all purposes.

BACKGROUND OF THE INVENTION

The genomes of all organisms undergo spontaneous mutation in the course of their continuing evolution generating variant forms of progenitor sequences (Gusella, *Ann. Rev. Biochem.* 55, 831-854 (1986)). The variant form may confer an evolutionary advantage or disadvantage relative to a progenitor form or may be neutral. In some instances, a variant form confers a lethal disadvantage and is not transmitted to subsequent generations of the organism. In other instances, a variant form confers an evolutionary advantage to the species and is eventually incorporated into the DNA of many or most members of the species and effectively becomes the progenitor form. In many instances, both progenitor and variant form(s) survive and co-exist in a species population. The coexistence of multiple forms of a sequence gives rise to polymorphisms.

Several different types of polymorphism have been reported. A restriction fragment length polymorphism (RFLP) means a variation in DNA sequence that alters the length of a restriction fragment as described in Botstein et al., *Am. J. Hum. Genet.* 32, 314-331 (1980). The restriction fragment length polymorphism may create or delete a restriction site, thus changing the length of the restriction fragment. RFLPs have been widely used in human and animal genetic analyses (see WO 90/13668; W090/11369; Donis-Keller, *Cell* 51, 319-337 (1987); Lander et al., *Genetics* 121, 85-99 (1989)). When a heritable trait can be linked to a particular RFLP, the presence of the RFLP in an individual can be used to predict the likelihood that the animal will also exhibit the trait.

Other polymorphisms take the form of short tandem repeats (STRs) that include tandem di-, tri- and tetra-nucleotide repeated motifs. These tandem repeats are also referred to as variable number tandem repeat (VNTR) polymorphisms. VNTRs have been used in identity and paternity analysis (US 5,075,217; Armour et al., *FEBS Lett.* 307, 113-115

(1992); Horn et al., WO 91/14003; Jeffreys, EP 370,719), and in a large number of genetic mapping studies.

Other polymorphisms take the form of single nucleotide variations between individuals of the same species. Such polymorphisms are far more frequent than RFLPs, STRs and VNTRs. Some single nucleotide polymorphisms occur in protein-coding sequences, in which case, one of the polymorphic forms may give rise to the expression of a defective or other variant protein and, potentially, a genetic disease. Examples of genes, in which polymorphisms within coding sequences give rise to genetic disease include β-globin (sickle cell anemia) and CFTR (cystic fibrosis). Other single nucleotide polymorphisms occur in noncoding regions. Some of these polymorphisms may also result in defective protein expression (e.g., as a result of defective splicing). Other single nucleotide polymorphisms have no phenotypic effects.

Single nucleotide polymorphisms can be used in the same manner as RFLPs, and VNTRs but offer several advantages. Single nucleotide polymorphisms occur with greater frequency and are spaced more uniformly throughout the genome than other forms of polymorphism. The greater frequency and uniformity of single nucleotide polymorphisms means that there is a greater probability that such a polymorphism will be found in close proximity to a genetic locus of interest than would be the case for other polymorphisms. Also, the different forms of characterized single nucleotide polymorphisms are often easier to distinguish that other types of polymorphism (e.g., by use of assays employing allele-specific hybridization probes or primers).

Despite the increased amount of nucleotide sequence data being generated in recent years, only a minute proportion of the total repository of polymorphisms in humans and other organisms has so far been identified. The paucity of polymorphisms hitherto identified is due to the large amount of work required for their detection by conventional methods. For example, a conventional approach to identifying polymorphisms might be to sequence the same stretch of oligonucleotides in a population of individuals by didoxy sequencing. In this type of approach, the amount of work increases in proportion to both the length of sequence and the number of individuals in a population and becomes impractical for large stretches of DNA or large numbers of persons.

SUMMARY OF THE INVENTION

The invention provides nucleic acid segments of between 10 and 100 bases from a fragment shown in Table 1, column 1 including a polymorphic site. Complements of these segments are also included. The segments can be DNA or RNA, and can be double- or single-stranded. Some segments are 10-20 or 10-50 bases long. Preferred segments include a diallelic polymorphic site. The base occupying the polymorphic site in the segments can be the reference (Table 1, column 3) or an alternative base (Table 1, column 5).

The invention further provides allele-specific oligonucleotides that hybridizes to a segment of a fragment shown in Table 1, column 8 or its complement. These oligonucleotides can be probes or primers. Also provided are isolated nucleic acids comprising a sequence of Table 1, column 8, or the complement thereto, in which the polymorphic site within the sequence is occupied by a base other than the reference base shown in Table 1, column 3.

The invention further provides a method of analyzing a nucleic acid from an individual. The method determines which base is present at any one of the polymorphic sites shown in Table 1. Optionally, a set of bases occupying a set of the polymorphic sites shown in Table 1 is determined. This type of analysis can be performed on a plurality of individuals who are tested for the presence of a disease phenotype. The presence or absence of disease phenotype can then be correlated with a base or set of bases present at the polymorphic sites in the individuals tested.

DEFINITIONS

An oligonucleotide can be DNA or RNA, and single- or double-stranded. Oligonucleotides can be naturally occurring or synthetic, but are typically prepared by synthetic means. Preferred oligonucleotides of the invention include segments of DNA, or their complements including any one of the polymorphic sites shown in Table 1. The segments are usually between 5 and 100 bases, and often between 5-10, 5-20, 10-20, 10-50, 20-50 or 20-100 bases. The polymorphic site can occur within any position of the segment. The segments can be from any of the allelic forms of DNA shown in Table 1.

Hybridization probes are oligonucleotides capable of binding in a base-specific manner to a complementary strand of nucleic acid. Such probes include peptide nucleic acids,

as described in Nielsen et al., *Science* 254, 1497-1500 (1991).

The term primer refers to a single-stranded oligonucleotide capable of acting as a point of initiation of template-directed DNA synthesis under appropriate conditions (*i.e.*, in the presence of four different nucleoside triphosphates and an agent for polymerization, such as, DNA or RNA polymerase or reverse transcriptase) in an appropriate buffer and at a suitable temperature. The appropriate length of a primer depends on the intended use of the primer but typically ranges from 15 to 30 nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. A primer need not reflect the exact sequence of the template but must be sufficiently complementary to hybridize with a template. The term primer site refers to the area of the target DNA to which a primer hybridizes. The term primer pair means a set of primers including a 5' upstream primer that hybridizes with the 5' end of the DNA sequence to be amplified and a 3', downstream primer that hybridizes with the complement of the 3' end of the sequence to be amplified.

Linkage describes the tendency of genes, alleles, loci or genetic markers to be inherited together as a result of their location on the same chromosome, and can be measured by percent recombination between the two genes, alleles, loci or genetic markers.

Polymorphism refers to the occurrence of two or more genetically determined alternative sequences or alleles in a population. A polymorphic marker or site is the locus at which divergence occurs. Preferred markers have at least two alleles, each occurring at frequency of greater than 1%, and more preferably greater than 10% or 20% of a selected population. A polymorphic locus may be as small as one base pair. Polymorphic markers include restriction fragment length polymorphisms, variable number of tandem repeats (VNTR's), hypervariable regions, minisatellites, dinucleotide repeats, trinucleotide repeats, tetranucleotide repeats, simple sequence repeats, and insertion elements such as Alu. The first identified allelic form is arbitrarily designated as a the reference form and other allelic forms are designated as alternative or variant alleles. The allelic form occurring most frequently in a selected population is sometimes referred to as the wildtype form. Diploid organisms may be homozygous or heterozygous for allelic forms. A diallelic polymorphism has two forms. A triallelic polymorphism has three forms.

A single nucleotide polymorphism occurs at a polymorphic site occupied by a single nucleotide, which is the site of variation between allelic sequences. The site is usually

preceded by and followed by highly conserved sequences of the allele (e.g., sequences that vary in less than 1/100 or 1/1000 members of the populations).

A single nucleotide polymorphism usually arises due to substitution of one nucleotide for another at the polymorphic site. A transition is the replacement of one purine by another purine or one pyrimidine by another pyrimidine. A transversion is the replacement of a purine by a pyrimidine or vice versa. Single nucleotide polymorphisms can also arise from a deletion of a nucleotide or an insertion of a nucleotide relative to a reference allele.

Hybridizations are usually performed under stringent conditions, for example, at a salt concentration of no more than 1 M and a temperature of at least 25 °C. For example, conditions of 5X SSPE (750 mM NaCl, 50 mM NaPhosphate, 5 mM EDTA, pH 7.4) and a temperature of 25-30°C are suitable for allele-specific probe hybridizations.

An isolated nucleic acid means an object species invention that is the predominant species present (*i.e.*, on a molar basis it is more abundant than any other individual species in the composition). Preferably, an isolated nucleic acid comprises at least about 50, 80 or 90 percent (on a molar basis) of all macromolecular species present. Most preferably, the object species is purified to essential homogeneity (contaminant species cannot be detected in the composition by conventional detection methods).

Linkage disequilibrium or allelic association means the preferential association of a particular allele or genetic marker with a specific allele, or genetic marker at a nearby chromosomal location more frequently than expected by chance for any particular allele frequency in the population. For example, if locus X has alleles a and b, which occur equally frequently, and linked locus Y has alleles c and d, which occur equally frequently, one would expect the combination ac to occur with a frequency of 0.25. If ac occurs more frequently, then alleles a and c are in linkage disequilibrium. Linkage disequilibrium may result from natural selection of certain combination of alleles or because an allele has been introduced into a population too recently to have reached equilibrium with linked alleles.

A marker in linkage disequilibrium can be particularly useful in detecting susceptibility to disease (or other phenotype) notwithstanding that the marker does not cause the disease. For example, a marker (X) that is not itself a causative element of a disease, but which is in linkage disequilibrium with a gene (including regulatory sequences) (Y) that is a causative element of a phenotype, can be used detected to indicate susceptibility to the disease in

circumstances in which the gene Y may not have been identified or may not be readily detectable.

The present invention includes the use of any of the polymorphic forms shown in Table 1 as a means to determine susceptibility to a phenotype resulting from an allele or marker in linkage disequilibrium with such polymorphic forms.

DESCRIPTION OF THE PRESENT INVENTION

I. Novel Polymorphisms of the Invention

The novel polymorphisms of the invention are listed in Table 1. The first column of the Table lists the names assigned to the fragments in which the polymorphisms occur. The fragments are all human genomic fragments. SGC, TIGR and WI respectively stand for Stanford Genome Center, The Institute for Genome Research and the Whitehead Institute. The sequence of one allelic form of each of the fragments (arbitrarily referred to as the prototypical or reference form) has been previously been determined. Many of these sequences are listed at http://www-genome.wi.mit.edu/); http://shgc.stanford.edu; or http://ww.tigr.org/. The Web sites also list primers for amplification of the fragments, and the genomic location of fragments. Some fragments are expressed sequence tags, and some are random genomic fragments. All information in the websites concerning the fragments listed in Table 1 is incorporated by reference in its entirety for all purposes.

The second column lists the position in the fragment in which a polymorphic site has been found. Positions are numbered consecutively with the first base of the fragment sequence as listed in one of the above databases being assigned the number one. The third column lists the base occupying the polymorphic site in the sequence in the data base. This base is arbitrarily designated the reference or prototypical form but is not necessarily the most frequently occurring form. The fifth column in the table lists the alternative base(s) at the polymorphic site. The eighth column of the Table lists about 15 bases of sequence on either side of the polymorphic site in each fragment. The indicated sequences can be either DNA or RNA. In the latter, the T's shown in the Table are replaced by U's. The base occupying the polymorphic site is indicated in EUPAC-IUB ambiguity code. The fourth and sixth columns of the table show the frequency with which reference and alternative alleles occur at a polymorphic site. The seventh column in the table indicates the population frequency of heterozygotes of the polymorphic site.

Table 1

Fragment	Position	Ref. Allele	Frequency (p)	Alt. Allele	Frequency (q)	Heterozygosity (h)	Sequence tag
SGC35469	118	A	0.75	C	0.25	0.38	TTAAGTGAGAMTCTTTTAAAC
SGC35512	34	T	0.5	C	0.5	0.5	AGAGCCGTCTYCTCAGGTTGC
SGC35512	50	G	0.31	C	0.69	0.43	GTTGCCTGTCSTCTCCTGGCC
SGC35594	74	C	0.63	G	0.37	0.47	GGCCGCATCCSTTAGTTTCCA
SGC35681	42	T	0.5	C	0.5	0.5	AGAGAAAAAAYCAACAGCAAA
SGC35681	56	A	0.56	C	0.44	0.49	CAGCAAACAAMACCACACAAA
SGC35683	34	T	0.75	G	0.25	0.38	CAATAAGCACKCATGACCTCA
TIGR-A003N21	49	C	0.94	A	0.06	0.12	GTGATTTGGTMAGCATATCTT
TIGR-A004S25	145	G	0.79	A	0.21	0.34	TGTACTTTGGRCTCCAGACTT
TIGR-A004V30	203	C	0.67	G	0.33	0.44	AGTAGAAAAGSCTTCTAGGTT
TIGR-A004W22	232	C	0.92	A	0.08	0.15	CCCCCGCCTAMCTGGAGATGT
TIGR-A004Z48	177	A	0.38	G	0.62	0.47	ACGCCACAGARTCCTCCAATT
TIGR-A005D24	123	A	0.94	G	0.06	0.12	ATAGAGAAATRAAAACCCAAT
TIGR-A005D24	138	C	0.75	T	0.25	0.38	CCCAATTTCTYTTTCACCATT
WI-10072	105	G	0.83	A	0.17	0.28	TATTTTTTGTRTGACTCCTAT
WI-10088	205	C	0.86	G	0.14	0.24	TTTAGACAGGSAGCAGAAGCA
WI-1017	93	G	0.57	A	0.43	0.49	ACCAGACAAGRGATGTAGATT
WI-1021	24	A	0.69	T	0.31	0.43	ATCAAAGCACWATCTGTGTTT
WI-1031	149	G	0.75	A	0.25	0.38	GATGCCAGCARCACAACACCC
WI-10396	72	C	0.29	A	0.71	0.41	TGGGAAGAGTMTGTGACTTTA
WI-10400	46	T	0.43	C	0.57	0.49	TAGAAAGTAAYTGCATTTCAG
WI-10400	165	A	0.86	T	0.14	0.24	CTCCCCACCCWAAAATAACGT
WI-10400	166	A	0.86	T	0.14	0.24	TCCCCACCCAWAAATAACGTA
WI-10400	189	A	0.43	G	0.57	0.49	TACCTATGTCRTGCCATGTAG
WI-10613	44	G	0.19	A	0.81	0.3	GAAACATACARTGTAATAGAA
WI-10613	172	A	0.06	C	0.94	0.12	ATTTTATTTGMGCCCTAGGAG
WI-10616	116	G	0.94	C	0.06	0.12	GTAGGTCCTGSTCTCCTATCA

SUBSTITUTE SHEET (RULE 26)

WI-10616	141	C	0.5	T	0.5	0.5	GCCACGTAGCYCTCCCTTCCC
WI-10656	59	T	0.5	G	0.5	0.5	TTTCTTTTTGKCTCTTAGAAT
WI-10673	94	C	0.69	G	0.31	0.43	ATGGAGGGGGSTGCAGGTTGG
WI-10681	41	A	0.58	T	0.42	0.49	GACCCCATTGWCTTACGCAAA
WI-10681	103	T	0.58	A	0.42	0.49	TAAAAAAGCCWAAAGACAGCC
WI-10685	25	A	0.86	G	0.14	0.24	TGGATAGGTCRACCGGCTGAA
WI-10744	61	G	0.33	C	0.67	0.44	ACAAAAGGACSAAAAACACTC
WI-10770	49	G	0.71	T	0.29	0.41	CTCCCTTTCTKCCTGGCCCTT
WI-10770	174	G	0.64	A	0.36	0.46	AGGACACTCARTTCACATGCC
WI-10809	33	C	0.71	T	0.29	0.41	AAACCATGAAYGGTATAAGGA
WI-10809	78	C	0.57	T	0.43	0.49	CCTCTCACCAYTTAGAAAAGG
WI-10826	132	A	0.71	C	0.29	0.41	AAGACCTGCAMCCCTGGCTTC
WI-10854	102	C	0.33	T	0.67	0.44	AGTTGAAACAYGAAGACGATA
WI-10854	152	G	0.33	T	0.67	0.44	CGAGGCAACAKGGAGAGGTAC
WI-10870	103	G	0.69	A	0.31	0.43	CCTACTTAGARCAGTGGAGTA
WI-11152	179	C	0.06	T	0.94	0.12	AGGCTTGTCAYCTGTCAGAAA
WI-11183	118	C	0.58	T	0.42	0.49	GTATTTTTGCYCTTGTCACTA
WI-11183	124	C	0.83	T	0.17	0.28	TTGCCCTTGTYACTAACATTT
WI-11183	192	T	0.83	C	0.17	0.28	GAGTTTTAAAYATTGGTATGT
WI-1126	97	T	0.13	C	0.87	0.22	TTTTCAAGATYCAATATATAT
WI-1126	230	T	0.63	C	0.37	0.47	GTAACTTTTTYGGACTTGTCT
WI-1795	47	T	0.38	C	0.62	0.47	ATGTCTGGGTYCTTCCAGACT
WI-1795	130	T	0.38	C	0.62	0.47	GAAAGAAAAGYCGTCTACCAT
WI-1819	51	C	0.94	T	0.06	0.12	CTTTCAGCACYTTCGTGGATC
WI-1936	117	T	0.58	C	0.42	0.49	TTGTATCACCYCTCCCGCAAC
WI-1968	167	A	0.75	G	0.25	0.38	TGGAAGTTGTRTGAACTTGAG
WI-2529	71	C	0.06	T	0.94	0.12	AGCCTCTCAAYTCTTAACTGC
WI-3429	62	C	0.5	T	0.5	0.5	GGGCTCCACAYAGCCCTCAGC
WI-3429	64	G	0.44	T	0.56	0.49	GCTCCACACAKCCCTCAGCCC
WI-3678	125	G	0.69	T	0.31	0.43	TGATGCACCTKCCTTTTGGAT
WI-4582	226	T	0.94	C	0.06	0.12	AAAATATGGTYCCTCCTTGCT
WI-4687	121	G	0.43	T	0.57	0.49	AAGGGCACTTKGCAGGAGTGT

SUBSTITUTE SHEET (RULE 26)

WI-4701	198	G	0.25	A	0.75	0.38	CCCAATTAGARCCATGTCATT
WI-4719	70	G	0.56	A	0.44	0.49	AGCGGATTATRTCTGACGCCA
WI-4767	50	A	0.33	G	0.67	0.44	CTTAGACTGARATTCATAAAG
WI-4767	173	C	0.83	A	0.17	0.28	AGGGATGACAMAAATCACTAA
WI-4823	164	C	0.5	A	0.5	0.5	ATTCCTAAAAMAAAGAAAAGT
WI-4860	72	A	0.71	G	0.29	0.41	TGCTTGATTTRGGAGATAAAA
WI-5222	52	G	0.29	C	0.71	0.41	CTCCATCCTASGATTCTGCCT
WI-5381	178	A	0.63	T	0.37	0.47	TTAGTTTTGTWTTACTAAAAC
WI-5385	110	G	0.67	A	0.33	0.44	CCAGGAATCGRCAATGCTAAT
WI-563	87	G	0.75	A	0.25	0.38	GGCCTCCCCTRCCCTGATCAT
WI-5696	61	C	0.07	A	0.93	0.13	CCTTAGTTTCMTAAAAGCCCC
WI-5760	187	G	0.5	A	0.5	0.5	TTAGATAAGCRTCCCACGAAA
WI-5801	48	A	0.25	G	0.75	0.38	GTGTCTTTGTRGAATTTGAAA
WI-5801	157	G	0.25	A	0.75	0.38	AGCCTGGGAARAGGGAATGAG
WI-5826	134	T	0.67	C	0.33	0.44	TATTCTTTAGYTTTCAAATTA
WI-5865	99	T	0.43	A	0.57	0.49	TATCAAAAATWAAACAAATAT
WI-5865	103	C	0.86	G	0.14	0.24	AAAAATTAAASAAATATTAAT
WI-5865	165	T	0.57	A	0.43	0.49	CAAGACACAGWCCAGTCTCCA
WI-5967	148	C	0.92	T	0.08	0.15	ATGCTTGGTAYTTGCTCTGTG
WI-5967	165	C	0.75	T	0.25	0.38	TGTGCCGTATYTGCTCCAATC
WI-6093	53	G	0.88	C	0.12	0.22	CTTTGGGCCASGTCTGTAATG
WI-6190	165	G	0.5	A	0.5	0.5	GAGGATCTTGRGAAGCAGCAG
WI-6213	164	C	0.94	G	0.06	0.12	TATACTATGTSATATAATAAT
WI-6238	175	G	0.56	A	0.44	0.49	TCTCAAAATTRGTTCCAGACT
WI-6275	148	G	0.43	C	0.57	0.49	GCTTGGGAAASGGAAGGAAAC
WI-6315	187	T	0.75	C	0.25	0.38	TTGCTGATAGYAGTGTCCTGG
WI-6315	193	C	0.94	T	0.06	0.12	ATAGTAGTGTYCTGGTTCTTC
WI-6554	195	C	0.86	G	0.14	0.24	GAGAGAAAACSCTGACTTTCA
WI-6644	134	T	0.92	A	0.08	0.15	CTCAAGCACAWACCCAAACTT
WI-6711	36	T	0.75	C	0.25	0.38	GACTCCAAAAYTGAATAAGTA

WI-6711	226	G	0.88	T	0.12	0.22	CACACCCACAKTGGCAACTAA
WI-6786	106	A	0.67	T	0.33	0.44	CTTTGGCGAAWGGATAAAGAA
WI-6786	111	A	0.5	T	0.5	0.5	GCGAAAGGATWAAGAAGTGAG
WI-6786	151	G	0.58	A	0.42	0.49	CCATTCTTCTRTGGGATAAGG
WI-6824	112	A	0.88	G	0.12	0.22	GTGCTGCCAARCACCTTAGAA
WI-6844	225	T	0.75	C	0.25	0.38	GTCTTGAGGTYATCATTATGA
WI-6905	215	T	0.75	A	0.25	0.38	ACATGAAAAAWAGAGCCTAAG
WI-6911	216	T	0.88	C	0.12	0.22	TTTACCACTTYCATGACATTG
WI-6962	78	A	0.63	G	0.37	0.47	GATCCAGAGARGACAAAGCTC
WI-7008	180	A	0.31	G	0.69	0.43	CTCTCAAAAGRAGAGTAGTTA
WI-7023	56	A	0.38	C	0.62	0.47	TTTGTGACAGMCCCTGCGTGC
WI-7023	206	C	0.31	A	0.69	0.43	ATTCAACACAMACACACATTC
WI-7038	31	G	0.69	A	0.31	0.43	GGACCTTGGCRCTCTCAGCTT
WI-7038	140	A	0.63	C	0.37	0.47	CCAGACAAGAMGACTGTCAGG
WI-7038	266	T	0.56	C	0.44	0.49	GAGACTTTTCYGCGTGATGGC
WI-7041	174	C	0.56	A	0.44	0.49	TCTGCCTCTCMCCACCTTCTT
WI-7069	93	G	0.13	A	0.87	0.22	TTAACAGAGTRTCAGATCTAT
WI-7070	226	C	0.94	T	0.06	0.12	ATGGTGCTTTYAGTTTAATGC
WI-7079	293	T	0.31	G	0.69	0.43	AGATGAAATTKATTTCCATCT
WI-7093	54	C	0.88	T	0.12	0.22	GCCCTTCCCTYGGCTCCCAGC
WI-7104	157	C	0.5	A	0.5	0.5	AGCATGAGGCMCAGCAAGAAG
WI-7104	249	C	0.56	T	0.44	0.49	AGCATCTTTGYTGGGCAGGGC
WI-7166	59	C	0.94	T	0.06	0.12	ATCAGTTCTAYGGATCATCAA
WI-7222	126	G	0.69	T	0.31	0.43	GGGGGATGGGKAATAAAGGAG
WI-7222	255	G	0.69	A	0.31	0.43	CATTTTCTCARTCATTTCCTT
WI-7224	134	T	0.94	C	0.06	0.12	TGTCAGCATTYATTAAAAAAC
WI-7227	24	A	0.88	G	0.12	0.22	CTCCTGGAGGRAGCCCAGGCA
WI-7227	93	G	0.5	T	0.5	0.5	TTTCAGACAAKCTTTAGAGAA
WI-7227	99	G	0.5	C	0.5	0.5	ACAAGCTTTASAGAAATGGAC
WI-7227	291	G	0.69	A	0.31	0.43	TAAGGGTTGARCAGTTAAAAC

WI-7259	189	T	0.44	C	0.56	0.49	CTGGCCACAGYTGGGGGAGCA
WI-7307	128	G	0.69	T	0.31	0.43	CCTCCCTCAGKAACTGGAGGA
WI-7310	64	T	0.13	A	0.87	0.22	ACAAGGAACCWCCGAAGAGGA
WI-7310	234	A	0.44	C	0.56	0.49	CCCCATCCCAMATGATCTTGA
WI-7313	256	C	0.25	T	0.75	0.38	TAGCGATGACYTCTTAATTAT
WI-7313	266	T	0.25	C	0.75	0.38	CTCTTAATTAYAATTTGATTT
WI-7322	275	A	0.5	G	0.5	0.5	ATAACAGAATRACTTGCCATC
WI-7330	207	C	0.5	T	0.5	0.5	AAAGTGAGAGYTGAAAAGAGA
WI-7381	54	C	0.25	G	0.75	0.38	GGGAAATCCCSCTTTCTTTCT
WI-7381	213	C	0.56	T	0.44	0.49	AAACGGCCTCYGGCTCTCAGA
WI-7416	137	G	0.06	T	0.94	0.12	TGGCAGTGCTKCTACTCCTCA
WI-7461	153	C	0.88	T	0.12	0.22	GACTGTGTCTYGTTCCCTGTT
WI-7587	28	C	0.56	T	0.44	0.49	AGGTAGCTCCYGAAGATCTGT
WI-7587	81	G	0.5	A	0.5	0.5	TCCCCTTCTGRATCTGAAAAG
WI-7587	133	A	0.19	T	0.81	0.3	CCTGAGGAAAWGGAATGAACC
WI-7676	139	C	0.56	T	0.44	0.49	GTGAAGGGGCYGGCTTCTCTT
WI-7676	309	A	0.5	C	0.5	0.5	GTGTCCTTGGMAAACTACCTA
WI-7685	46	T	0.13	C	0.87	0.22	TTTTGGGCTCYTTTTTCTCCC
WI-7718	42	A	0.44	C	0.56	0.49	TTACTCAAGCMGTTACTCCCT
WI-7718	222	C	0.31	T	0.69	0.43	TTACAAAGAAYCATGCAGGAA
WI-7718	248	A	0.5	G	0.5	0.5	ACTATGTATTRATTTAGAATG
WI-7719	163	A	0.63	G	0.37	0.47	ACAGTTATCCRTTAGATCAAG
WI-7719	281	T	0.19	C	0.81	0.3	ATCTAGAATCYCTTTATGTTC
WI-7721	145	A	0.75	C	0.25	0.38	CTGTCTCTGCMTCTGACTCTC
WI-7805	101	A	0.25	G	0.75	0.38	GAATATGTGTRTGTTAAAGGA
WI-7842	57	T	0.58	C	0.42	0.49	TCCCATTCTGYGTATGAGTCC
WI-7850	57	G	0.69	A	0.31	0.43	CTGCCTCTGGRCTCATGTATC
WI-7860	50	C	0.75	T	0.25	0.38	CCTCTCCCCAYTGGGGAGAGA
WI-7878	51	C	0.25	G	0.75	0.38	TGATGGCCTGSTGGTTGATAA
WI-7878	162	A	0.19	G	0.81	0.3	GGAGGAGCTGRGTGTGATGAA

WI-7928	101	T	0.14	G	0.86	0.24	TCAAAATTCAKACAAGAGGAA
WI-7933	96	G	0.75	A	0.25	0.38	TTGGCCAGGGRCCTCGTATCC
WI-7936	131	T	0.56	A	0.44	0.49	TACACCAAACWACTGAATGAA
WI-7944	99	T	0.19	C	0.81	0.3	GACTTTCATGYAGCCCAAAGT
WI-8007	242	C	0.92	A	0.08	0.15	ACTGTTGGACMAGCTGCTGGA
WI-8010	247	G	0.75	T	0.25	0.38	AGTGGTGGGGKCTTCCACGTG
WI-8039	87	T	0.94	C	0.06	0.12	TTGTTTCAGTYAAATATGTAT
WI-8039	97	T	0.06	C	0.94	0.12	TAAATATGTAYGTGTCCGTGC
WI-8044	107	C	0.58	A	0.42	0.49	GGTTTCTCCCMAGTATGGATT
WI-8053	242	T	0.08	A	0.92	0.15	ACTTATATAAWTTCAGAACTA
WI-8054	131	C	0.63	G	0.37	0.47	CAAGCCTTAGSACAATCTTCT
WI-8054	148	T	0.56	C	0.44	0.49	TTCTTTGTAGYTTTAGCCTTT
WI-8054	237	G	0.5	T	0.5	0.5	GGCGTACAGAKAATCCTTGCC
WI-8057	87	T	0.57	A	0.43	0.49	AAAAGGACAGWGATGGACAGC
WI-8170	204	T	0.88	A	0.12	0.22	CAATCAGAAAWAAAGGTAAAA
WI-8170	259	G	0.56	A	0.44	0.49	ACAAGAAGCARGCACTTAAAT
WI-8456	93	G	0.38	C	0.62	0.47	GGATGTCACASTTATGTCAAG
WI-8496	41	G	0.79	A	0.21	0.34	GAATGGTAATRTTGTATCAGT
WI-8496	157	A	0.79	G	0.21	0.34	TGCCAATGCARTTAGTATATA
WI-867	119	G	0.56	A	0.44	0.49	TTTCATCTCCRTTTGTGTGTT
WI-931	31	A	0.5	G	0.5	0.5	CGGAAGCCACRGCCACTAGCC
WI-931	191	C	0.5	A	0.5	0.5	CAAAAAAGCCMCGAGCCTGGT
WI-9443	211	G	0.81	A	0.19	0.3	CTGACGAGACRCAGAGACCTT
WI-9448	184	G	0.31	A	0.69	0.43	CTGGCACCACRCACTGGTTTC
WI-9484	178	G	0.92	A	0.08	0.15	GCCAGACAGGRAGGAATTCAA
WI-9617	37	G	0.88	T	0.12	0.22	ACACGCCGTGKTGGCACAGTC
WI-9651	105	A	0.56	T	0.44	0.49	TCGTCCTTCAWGGGGCAGCTT
WI-9651	139	T	0.88	C	0.12	0.22	TAGACACCTCYACAGGTACAG
WI-9657	121	T	0.67	G	0.33	0.44	CAAAATAAAGKATAATTCTTT
WI-9667	68	G	0.81	C	0.19	0.3	TTGTATCATGSTTATCACTGG

SUBSTITUTE SHEET (RULE 26)

WI-9667	82	C	0.75	T	0.25	0.38	TCACTGGACAYAGCCACCTCC
WI-9702	179	C	0.56	T	0.44	0.49	CAGTTTTATTYTAACTTTAAT
WI-9702	344	C	0.5	T	0.5	0.5	AAGACTGGAGYGCTCAGCCTG
WI-9702	345	G	0.38	A	0.62	0.47	AGACTGGAGCRCTCAGCCTGC
WI-9705	111	C	0.5	A	0.5	0.5	TTCGGCTGCCMAAAATTGTTA
WI-9711	390	C	0.5	A	0.5	0.5	GGCATAAGTGMAGGAAAGAGA
WI-9711	423	T	0.69	A	0.31	0.43	AGGAAAAAAAWGTTATCTGCT
WI-9716	221	G	0.81	A	0.19	0.3	AATTCTAGAARAAAACACCTA
WI-9760	49	C	0.86	T	0.14	0.24	CTCTCTTTACYAAGTGTTACT
WI-9814	104	C	0.92	T	0.08	0.15	GCTGCTATCTYTTCTCCTTCA
WI-9823	97	C	0.57	T	0.43	0.49	GTGAAATTTCYGGGGCATGGG
WI-9825	123	A	0.94	T	0.06	0.12	TCAGGGTGCTWGAGGATTAGT
WI-9826	125	A	0.5	T	0.5	0.5	AGAGGCTGTTWTGGCCTTCAA
WI-9826	127	G	0.5	A	0.5	0.5	AGGCTGTTATRGCCTTCAAAG
WI-9855	31	A	0.17	C	0.83	0.28	GAAACTGTAGMAAATTCTTTT
WI-9891	39	T	0.44	C	0.56	0.49	ACTGCCTCCTYAGTGAGCCTG
WI-991	37	A	0.63	T	0.37	0.47	TTCTGTACATWCATTATTGTA
WI-9975	126	C	0.88	T	0.12	0.22	GCCTAGAATAYAGTGGGTCCC
WI-9983	146	C	0.69	T	0.31	0.43	AGCATTATGAYAGACACAAAG
WI-9986	42	T	0.75	C	0.25	0.38	ACAATTTGAAYGTACCCCAGG
WI-14263	49	T	0.63	C	0.38	0.47	AAAAGGCATATTCAAYTGTCCCATACTAATT
WI-14267	28	T	0.94	C	0.06	0.12	ATTAGGAAGGGAGCAYTGAAATGGGAAGGGG
WI-14284	55	C	0.94	T	0.06	0.12	TTTAGTGCAAAAACAYTATGCCATGCGGGAA
WI-14288	85	G	0.38	C	0.63	0.47	CTGCTATTCCCAGATSAAGATTTGGTGGAAG
WI-14297	86	A	0.81	T	0.19	0.30	GGTACTTTTTCCAAGWAAAATGTTTCTGAAT
WI-14319	83	C	0.19	T	0.81	0.30	AGGCACAAAGCTAAGYACATGCAACAATATA
WI-14323	78	T	0.75	C	0.25	0.38	AAGAATCAAACATCAYTCTGGACCATGGGAA
WI-14323	86	C	0.94	A	0.06	0.12	AACATCATTCTGGACMATGGGAACCTTGAAA
WI-14339	102	T	0.81	G	0.19	0.30	ACAGTACATGATTACKCGGTTTCCAGAAATC
WI-14372	86	A	0.94	G	0.06	0.12	TCAAATAAATAGGGARTTCTCTTTAAATAAC

WI-14373	95	A	0.94	G	0.06	0.12	CCCTGGACGAAACCARCACATATACAATCAT
WI-14379	102	C	0.44	T	0.56	0.49	GGGTTATGTCACACCYTGTCAACCTCAAAAC
WI-14408	60	T	0.69	A	0.31	0.43	CACTATTACAGGCTGWAAAGTAACAAATGAG
WI-14482	17	G	0.88	A	0.13	0.22	AGAACCAATTAATAARAATCTGCAAGTTTTC
WI-14492	92	A	0.69	T	0.31	0.43	AAATTACTAAATTAAWGTCTTAAAAGAAAAT
WI-14510	104	A	0.25	T	0.75	0.38	TATGCATAACAAAATWTGCCAGTTTAACCAT
WI-14528	62	T	0.75	G	0.25	0.38	CTGGATGGTATAAATKTTGAATTATAAATTT
WI-14546	95	C	0.81	A	0.19	0.30	ATAGTAGAGGACTCAMCCTGCACGTGCACCT
WI-14580	100	G	0.69	A	0.31	0.43	CCCATCTGTCTTGCARGGAGGGATCTTGGTC
WI-14631	82	G	0.94	A	0.06	0.12	TCTGTCTTCTTTAACRTGCCTGGTTCCCTCT
WI-14635	22	G	0.94	A	0.06	0.12	AGATACAGAGCTGTCRTCTTGAAGACCACCA
WI-14651	49	C	0.88	G	0.13	0.22	CTCATTTAAAATTGTSAAATAAGTCAGAAAA
WI-14666	105	T	0.63	A	0.38	0.47	AGCTAATGTATTAAAWAACCATGAAAAGAAA
WI-14683	91	A	0.88	T	0.13	0.22	TAGTATCTAAAAACAWCAAAAAAAACACTGG
WI-14712	38	T	0.63	A	0.38	0.47	TCCAAGTACAAATCAWCTCACAATACCATAT
WI-14733	98	G	0.50	A	0.50	0.50	GACAGATATTCTGCARAATAAATGGCCTGAC
WI-14759	73	T	0.56	C	0.44	0.49	GTTTGACTTGTGCGGYGTACTCAAATGGGGG
WI-14808	52	T	0.69	A	0.31	0.43	ACCACACTACCCTGTWAAAATCTTAACATTG
WI-14816	29	A	0.69	T	0.31	0.43	GAGTCAGCATTTATTWAAAAACTGGACACGC
WI-14836	28	T	0.94	C	0.06	0.12	AGAGGACAGAGTGTTYGTTGATTTTTCGTTT
WI-14856	60	A	0.88	T	0.13	0.22	CGGAAAATACTTAATWTAAAGTTTGTAAAAA
WI-14863	61	G	0.94	A	0.06	0.12	AATATTTTTTGTCTGRAGTTAATAAAGTTAA
WI-14867	46	T	0.56	C	0.44	0.49	CAAGGCTCTCTAACAYGAGTGTCTGCAGCCC
WI-14898	50	A	0.88	C	0.13	0.22	GAAGAGTTGTCTCATMAGGTGCCACTAAGGA
WI-14898	79	A	0.88	C	0.13	0.22	GAAAACTTTCTCCATMAAGCTGCCTGCTGTG
WI-14907	48	G	0.81	A	0.19	0.30	ACATTGGACTCTGACRATTCCCCTTGCAGCA
WI-14911	52	G	0.38	A	0.63	0.47	ATTCAGTTCCTGGTCRAAGGTCCTTTTCCTG
WI-14913	88	C	0.88	A	0.13	0.22	ATAGTAGAGGACTCAMCCTGCACGTGCACCT
WI-14914	66	G	0.63	C	0.38	0.47	CAGTTTTCTCTAGCASGAATTTATTGTCCTG
WI-14926	49	T	0.94	C	0.06	0.12	TGGGCACTTAGCGAAYACTTGTGGACCACAA

SUBSTITUTE SHEET (RULE 26)

WI-14930	55	C	0.81	T	0.19	0.30		GAGTCCCTCATGGATYGCGGTATTGGTTGGT
WI-14946	47	T	0.94	C	0.06	0.12		CCCCCAGACATAACAYCTCTAAATCATCCTC
WI-14948	56	T	0.13	C	0.88	0.22		CTGCTAACTTGTCAGYTCCAACAACTGATGT
WI-14958	83	A	0.75	G	0.25	0.38		CTTTCTTTTCAAGGGRAAAAAACCCAAATGA
WI-14976	35	C	0.44	T	0.56	0.49		TTGCTTCGTTCAAAGYGCTTAGAATGGAAGA
WI-14981	31	G	0.38	T	0.63	0.47		GTTTATTGGATTTTTKTTTATGCTAAGTATT
WI-14992	80	C	0.25	T	0.75	0.38		TAAATGAAGCTGCAGYAGGAAAGCTGAGCAC
WI-15000	90	G	0.88	A	0.13	0.22		CAGACTGTCTAAGTARTGAAGTTTGTGCAGA
WI-15002	72	T	0.94	A	0.06	0.12		GCCTTCTTGATTTCCWTTCAGTTTAGGCCTC
WI-15012	59	G	0.56	T	0.44	0.49		TTTCATTGAAGCTTTKTACCTTACTATACTC
WI-15069	81	T	0.94	C	0.06	0.12		ACGCACTAAAAAAAAAYGTGTGCTTGCTGCTG
WI-15100	74	G	0.94	A	0.06	0.12		GACTGGAGTGAGAACRGGTTCCACCACCAAG
WI-15116	96	C	0.81	T	0.19	0.30		CCCTAGTTGCAGTAAYGTGTCATAATAAATA
WI-15123	55	C	0.63	T	0.38	0.47		CAGATAAATAGGATGYGTCTGTTTGCCCTTA
WI-15152	51	G	0.94	A	0.06	0.12		CTATGTAACTACACARTATGCACACCACAGC
WI-15153	40	A	0.81	G	0.19	0.30		TATGTTGGCATTGCARAGACACTGCACTTAT
WI-15182	49	C	0.88	A	0.13	0.22		AACCAGGGCAAAATAMTGCTGGATTAACCCA
WI-15198	38	T	0.38	C	0.63	0.47		GCCCTTGGCACTATGYCTACTCTGCCTGACG
WI-15215	84	G	0.44	C	0.56	0.49		TTAGAATCAAATGGGSTGACTTTTTCCCCTG
WI-15225	80	C	0.75	T	0.25	0.38		ACCTAGAAAGCAAACYGGAGTGATTATGCCA
WI-15239	57	T	0.56	C	0.44	0.49		AATAAACACCATCATYCCTGAGTCCACAGAT
WI-15249	34	T	0.81	C	0.19	0.30		ACAAAGTTCTAACTTYTTGTTAAAAATCTCT
WI-15260	75	G	0.63	A	0.38	0.47		GAAGCTAATCATGGARGCAAGCTCCCTGGAG
WI-15288	108	C	0.63	G	0.38	0.47		AGGATTCCCTCTCTCSTCCAAGGGAAAGAAG
WI-15295	27	G	0.63	C	0.38	0.47		GAATGTATTCCTGATSTTTTCCTTTGCCAAC
WI-15325	39	T	0.13	C	0.88	0.22		ATGTGGCTGGGAGGCYTCACAATCATGGTGG
WI-15347	74	C	0.81	T	0.19	0.30		GAAAAGAACAAATTTYCAAAGACTTGGGGGA
WI-15353	37	G	0.94	A	0.06	0.12		CAATGTGGTGAAAACRTCTTAATTCAGGACA
WI-15361	101	A	0.56	G	0.44	0.49		GAACTCAAGTCATCARTTTTAGGCACAAAGG
WI-15389	33	G	0.69	A	0.31	0.43		AGCTTGCTTTTTGTCRTTTGGAAGACTACCA

WI-15389	104	G	0.81	A	0.19	0.30	AAACATCTGCGAAAARAAGTGTGGGAATCAC
WI-15407	92	A	0.56	G	0.44	0.49	AAGGATTAAGTTTAARCCACACTACCAAAAG
WI-15488	69	C	0.31	T	0.69	0.43	CAGCCAGATATCAACYGTTACAGAAATGAAA
WI-15625	40	C	0.38	T	0.63	0.47	AAAAGGCATATTCAAYTGTCCCATACTAATT
WI-15702	48	G	0.63	C	0.38	0.47	AAGGCTTTCAAAAAGSGGGGTAAAGGGGTGA
WI-15702	90	C	0.56	T	0.44	0.49	GAGAGAAACTGTAACYCTGTAAACAATACTA
WI-15702	101	T	0.69	C	0.31	0.43	TAACCCTGTAAACAAYACTAATGGGTTCTTT
WI-15702	107	T	0.31	C	0.69	0.43	TGTAAACAATACTAAYGGGTTCTTTGAACAA
WI-15705	50	A	0.13	G	0.88	0.22	ATTTTAGACTGAATCRTTCTAGAGTATTTGA
WI-15719	69	A	0.63	C	0.38	0.47	TTTCATCCATTCAGCMAATTTAAAACTCTTG
WI-15729	35	A	0.56	G	0.44	0.49	CCATGTGTAGACTGCRGGCACTTTAGAAAGA
WI-15736	27	G	0.81	T	0.19	0.30	CATTAAACTTGCACAKTAGCAAAAAAAATCA
WI-15747	88	T	0.69	C	0.31	0.43	ACTAATTTAGTGTTTYTTTAAATTATATGAA
WI-15801	24	G	0.81	A	0.19	0.30	CCAAGAATGGGAAGCRCATTTTCATTGGCTT
WI-15801	81	T	0.63	G	0.38	0.47	TAGCTGCAGTAATACKGCATCCCATCCACTC
WI-15809	77	T	0.38	G	0.63	0.47	TCTGTTGTAAATGCCKTTTACAAACATTGAA
WI-15843	62	C	0.25	T	0.75	0.38	CCAAGAAGCCTTCAGYAGAGCAAGTCTGAGC
WI-15868	21	G	0.69	C	0.31	0.43	ATGCAATGAATAAAASGGCAGAAAATTCAGA
WI-15892	123	A	0.94	T	0.06	0.12	AACCAAGAGAAGGAAWGGAATCAACTCCACA
WI-15937	24	A	0.75	G	0.25	0.38	CTGCTGTATTTAAAARACAAGCGTCTGGATC
WI-15944	24	A	0.88	C	0.13	0.22	AACGTATTTCCTCCAMACACCGTAGAAACTT
WI-15953	26	T	0.31	G	0.69	0.43	TGTCCTTCACATCATKTATATTGTATTGCAC
WI-15953	59	C	0.56	T	0.44	0.49	AAACTTTTTTAACTCYGTCAAAAACAACAAG
WI-15964	99	T	0.88	A	0.13	0.22	CTGTCCCTGGAGGTAWGCAAGAGGGTGGAGA
WI-15986	60	T	0.69	G	0.31	0.43	TGTGGGTTTTTTTTTKTTACATTTTCTTTTA
WI-15987	32	C	0.38	T	0.63	0.47	TTAAAGGGGTCCCAAYGAGGTTGGTAGTGCC
WI-15987	80	A	0.88	G	0.13	0.22	ACTAAGAAGATGGTCRTCTATGAACCAAGCT
WI-16002	59	T	0.25	C	0.75	0.38	ATCATGAGAATTTCAYGTTAAAAGTCAAAGA
WI-16083	89	C	0.88	T	0.13	0.22	AAACATATCAAGGATYGGGCTGGAATCTTTT
WI-16100	52	A	0.69	G	0.31	0.43	TTTCCTACACTTGACRGTAATATACTGTTTT

SUBSTITUTE SHEET (RULE 26)

WI-16156	97	A	0.56	C	0.44	0.49	TTAACCCAGAGTCGCMTCTCTTCAAAATGCA
WI-16163	35	C	0.50	T	0.50	0.50	ATGCAATTGAAATAAYATTGTAAGTTAATGT
WI-16167	58	T	0.88	C	0.13	0.22	TTTCTGATATACATTYCATCTTATTCACCAC
WI-1011	70	G	0.86	C	0.14	0.24	AAGTTTTTGTCTCCASAGAAGTCATTTTGTA
WI-1172	17	C	0.57	A	0.43	0.49	AACGTGTGGTTAAAAMTAGGCAATTGGTTAA
WI-1172	179	C	0.43	T	0.57	0.49	ATGGCTGATACCAAGYCTGCAGTGAAAAATG
WI-1177	35	G	0.14	C	0.86	0.24	AAAAAAATGAAAGAASAAGAAAAAAAGAGTC
WI-1231	126	T	0.71	C	0.29	0.41	ATTCTCCTTCTTTCAYTAATTTTCTTTCACG
WI-1231	141	G	0.71	A	0.29	0.41	TTAATTTTCTTTCACRTTATTCCCTCACCCT
WI-1319	40	A	0.50	T	0.50	0.50	CATAGTTTATTCTTTWACCATAGGGGTGTGT
WI-1356	123	T	0.79	C	0.21	0.34	CAAGAAAAAAAAAGCCYGTACATGTTTGGTAC
WI-472	114	G	0.86	C	0.14	0.24	TATACAACAGAAAAAGSGGGCTGGAAAAGAAA
WI-478	46	C	0.64	T	0.36	0.46	TACTCTATTTTGTTCYAGCCACCTGTGGCAT
WI-533	29	T	0.36	C	0.64	0.46	AGTACCTTTCTAACTYATAAGATTGTGTAGA
WI-601	74	C	0.07	T	0.93	0.13	AAAGATGGTAGTGAGYGAACAGAAGAGGTTT
WI-601	112	T	0.64	A	0.36	0.46	TCCTAAACTGAGTACWCAAAAACGAGCAGGT
WI-863	107	A	0.64	G	0.36	0.46	TTCACAACCTCACCARACTTGGCTTACCGGG
WI-919	36	G	0.64	A	0.36	0.46	TTAATCAACCTAGCCRGCTGTCATGTGGGAT
WI-1736	175	C	0.92	T	0.08	0.15	TCCATCTGTCTTCCAYAGAGATCTAGGGTGT
WI-1754	177	G	0.33	A	0.67	0.44	CTTAAAGAGATAGTCRCCAGAGGCAATTCGA
WI-1775	47	C	0.83	T	0.17	0.28	ATGGTCTTTTCTCTGYTTTACATCATTGTCA
WI-1851	136	G	0.83	A	0.17	0.28	TATTAACATGGTACARACAACTTCAGTTTAA
WI-1949	86	T	0.42	G	0.58	0.49	TGAGATGCTCTGAGTKCAAGGCTGCTGACAT
WI-1949	160	T	0.50	C	0.50	0.50	ATGAATGCCATAATCYCTGTGTTTTTTGTCC
WI-1965	105	G	0.67	C	0.33	0.44	AGGAAGTGTTTAAAGSAGAGAGATGACCCAT
WI-2020	145	C	0.92	A	0.08	0.15	TGGGTCAACTATGATMCCAAAACAGCAGTGT
WI-2028	176	T	0.17	C	0.83	0.28	GTTCCCTGTCTCATCYTTCTAGGTAATTTGA
WI-2033	183	T	0.25	C	0.75	0.38	AGAACTAATCCCTCAYGGAGAACGTGGAACC
WI-2034	150	T	0.42	C	0.58	0.49	CAGTGCACCAAGGACYGGACCTGCACTCTAT
WI-2038	155	C	0.83	T	0.17	0.28	ATTTCTATTTTGATAYTGATGTTTCTTTCAA

WI-2287	24	T	0.92	C	0.08	0.15	TCTGTGGTCCCCTTTAYAAAGCCTCTTGCATC
WI-2296	81	A	0.50	G	0.50	0.50	ATTCTTTGCTCTGACRCCAGTTAGCTGTGTG
WI-2300	77	G	0.33	T	0.67	0.44	AGAAGCCAGTCATACKTGCTTTAAAATTGAC
WI-2371	55	G	0.69	T	0.31	0.43	TTCTTCCCAGCTTCTKGTGGTGGCTGTCAAT
WI-2395	122	A	0.69	C	0.31	0.43	AAAATTACTATCCAAMCTGAATTCAGAATAA
WI-2437	128	G	0.06	A	0.94	0.12	CCAAAAAATTCCCAATRCTCTAAATAGATGGA
WI-2437	179	G	0.94	A	0.06	0.12	CAAGAGGCAATCGACRAACATCACAGTGGGC
WI-2437	192	G	0.94	A	0.06	0.12	ACGAACATCACAGTGRGCTGTGGTGCCAAGG
WI-2440	71	G	0.88	A	0.13	0.22	ATTTAATTTTAGTTGRGTGAGACCAATAGCA
WI-2572	61	C	0.94	T	0.06	0.12	AACACTTCTCCCACAYACAAAGTTAACACTT
WI-2616	125	T	0.13	C	0.88	0.22	CAAGAATTGATCCTAYACTGGGACTACAGCC
WI-2625	98	G	0.00		0.00	0.00	AAGGCTTATTTAGGA_CAAATTGATGATACT
WI-2716	23	T	0.88	C	0.13	0.22	ATCCAGAAAAACAGCYGAATGACACAACAAGAG
WI-2886	46	C	0.81	A	0.19	0.30	GTCTGGGGGAGAAGAMAACGAGATAAAGCAT
WI-2906	50	A	0.25	C	0.75	0.38	CTTCATTCTTGCTGGMACTTTGCCTGGAATG
WI-2906	77	T	0.31	A	0.69	0.43	AATGCTCTTTCCCTCWGAGCTTTGCTTGGCT
WI-2924	54	G	0.75	A	0.25	0.38	GTCTTCTCTTATAGGRACCCTGTGATTACAC
WI-2939	72	G	0.63	T	0.38	0.47	TGTCTCAGTGCCTTTKCAAGACCTTCCCTCA
WI-3000	62	G	0.38	A	0.63	0.47	AAACACAGAGACCCCRTGAGTCTTAGTCAAT
WI-3167	37	T	0.88	A	0.13	0.22	AGATCTATTAGATTCWCACCCATCTCAAAAC
WI-3203	99	G	0.63	A	0.38	0.47	TATGCCGCAGACGAGRCCACACAAGGCAATA
WI-3208	140	G	0.69	A	0.31	0.43	GTGGGCAGATAAAGARCCAAGCCCTAGTTTG
WI-3275	157	C	0.94	G	0.06	0.12	CAGAACTATTTCTCASTAAGAATCTTAAGTT
WI-3402	55	G	0.50	A	0.50	0.50	TTGATTTCCTTACATRCAAATGCTCCTTTTT
WI-3416	33	C	0.69	T	0.31	0.43	TAGCATTCAGAAGTCYCTCTTAGAGGTAGTT
WI-3453	70	C	0.19	T	0.81	0.30	GGCCCATCAGAGAATYGAAGTCATGGGGAAA
WI-3473	101	A	0.88	G	0.13	0.22	TTTTAGCCCTAGGGARTAGAAAATGTTGGTG
WI-3474	90	A	0.38	G	0.63	0.47	CCCTAATTTTAGCACRGTATTTTAATGAGGT
WI-3474	109	G	0.94	A	0.06	0.12	TTTTAATGAGGTGGTRTGGGAGAAAATTGAT
WI-3502	79	C	0.56	T	0.44	0.49	GGTTTCTGGATGTCTYTGAGGACAGGGTCAC

WI-3600	78	T	0.88	G	0.13	0.22	CCCCTGATAGTTCTGKGAGCCACCTAAACTC
WI-3600	146	G	0.56	C	0.44	0.49	TGGATATAAACATCTSATGGAAGGCTGCACT
WI-3687	67	A	0.94	C	0.06	0.12	AATATGACATAAAATMAAAAACTACTATAGT
WI-3735	72	T	0.63	C	0.38	0.47	TATCAAATGAAAAACYACACCGGTTCAATGA
WI-3746	116	G	0.94	A	0.06	0.12	CATCTCTGTCTCTGCRGCCCCAGGATAAAGC
WI-3867	49	T	0.69	C	0.31	0.43	TAGTCTTCCTGACAAYCGGATGTACCTAGTA
WI-3898	25	A	0.71	C	0.29	0.41	TGTCTTTAGAAGCAGMGGAGAGGACACCGAC
WI-3901	114	A	0.07	G	0.93	0.13	TCACCTGACAAGTGGRTATCATGTGCTACAC
WI-3914	99	C	0.71	T	0.29	0.41	CTCAAGACTCACAGCYACCATCCTTCATTGC
WI-4019	33	G	0.36	A	0.64	0.46	CGTCCTATGAATCATRCATTTGTTCCTGTTA
WI-4091	84	A	0.71	T	0.29	0.41	CTTAGTCATTGCATGWTGTATAACAATATTG
WI-4160	117	A	0.86	G	0.14	0.24	ACAATATCAACAGAARGGCTATATTAGAAAA
WI-4168	32	A	0.86	G	0.14	0.24	AAATTGATACAAACARTCTGAAAATCTGTTT
WI-4177	68	T	0.64	C	0.36	0.46	TACCTATTATATTTAYCATCATGATTTGCTG
WI-4199	51	A	0.43	C	0.57	0.49	AGTCAATATAAAAAAMCACACATATTGTTAT
WI-4250	94	G	0.36	T	0.64	0.46	GTCTTGTGAAACAGGKGTGGGAAGGATCCTG
WI-4250	117	A	0.57	G	0.43	0.49	GGATCCTGTAAAAGGRTAAATATTGTTTTCC
WI-4255	68	G	0.86	C	0.14	0.24	GCTCCCCCATCACCTSCCTTACACAACTTGA
WI-4256	57	C	0.93	T	0.07	0.13	AGAGGCAAAATCTGGYCTCACCATTGGAAAA
WI-4325	58	C	0.93	T	0.07	0.13	GTACATGGGCAGGACYGGAAATGGGATGCTA
WI-4325	71	C	0.57	T	0.43	0.49	ACCGGAAATGGGATGYTACTATAGATAATCT
WI-4347	158	A	0.07	G	0.93	0.13	TATCTGTTCAGGCCCRGAATCGTCACGGCTC
WI-4360	93	C	0.63	T	0.38	0.47	GTATTTTCCAAATAAYAAAATGCCTCTGAAA
WI-4448	112	T	0.63	G	0.38	0.47	AGATGGGGTATATAAKAAAGAACCATGTAAA
WI-4456	49	C	0.69	T	0.31	0.43	GAAAATTATAGTTCCYCAAGTTCATGCATAA
WI-4461	49	A	0.50	G	0.50	0.50	TAAAATTATCCTTCCRTGAAATTGGTGAAAG
WI-4465	41	A	0.75	G	0.25	0.38	AGACAACACGAAAGTRTATAAAGAAAACAGT
WI-4465	75	G	0.75	A	0.25	0.38	TAATCTTTCACCTTTRTATTTCTCTTCTACC
WI-4529	64	T	0.44	C	0.56	0.49	ATCATTCTGAAGATGYGAGTTCTTCTTTTAT
WI-4540	110	A	0.88	G	0.13	0.22	CACCATGTGGCATCCRTGCATGGCTGCATTG

WI-4596	69	T	0.25	A	0.75	0.38	AGAAAGCACTGTGACWCATTATTAGGCCCAT
WI-4606	61	A	0.56	G	0.44	0.49	AGAAAATTATGCCTARCCAAGTAGACAACTT
WI-4649	50	C	0.44	T	0.56	0.49	CATTCTTTCCGAATGYGATGATTTCTTGTAA
WI-4650	148	A	0.13	G	0.88	0.22	TCTTATATTGCTTTTRCCAAATCCAGTTTAA
WI-4677	82	T	0.69	C	0.31	0.43	GAGTTGAAATAAATGYAAGTTGAATAATGAC
WI-4698	135	C	0.94	G	0.06	0.12	GGAAGAAAACTTCAASTTCGAGAAGGCTTAG
WI-4722	88	G	0.81	A	0.19	0.30	TATGGAACACCACACRCAACTGAATGCAGAT
WI-4745	131	T	0.75	C	0.25	0.38	TACTTTCTACTCTGAYAGGCAGACTTATATG
WI-4782	113	C	0.63	T	0.38	0.47	ATAACTAGAAAATGCYGAACAGAAAAATAAC
WI-4788	65	A	0.75	G	0.25	0.38	ATCTTGCTAAGTTCCRTGAAAAAAAATTATG
WI-4818	43	A	0.38	G	0.63	0.47	GACTAGGTTATGTCCRCACATGAATAAACAA
WI-4818	121	G	0.56	T	0.44	0.49	TAATGGGGCCCTGTTKCTCTGGCATACATAT
WI-4888	56	G	0.81	A	0.19	0.30	GAAAAGATAACAAGARATGAATAAATGAGGT
WI-4897	93	A	0.94	G	0.06	0.12	AAAATAAGCGCTTGGRGATAAACACATCTTC
WI-5163	24	C	0.38	T	0.63	0.47	CACTGGTCTGCCTGTYGGTCTGTTCCTGTGT
WI-5204	54	C	0.94	T	0.06	0.12	TTGGGTTTTGAAGAAYGAAGAAAAATGGAAA
WI-5215	70	A	0.81	G	0.19	0.30	CAGACTCAAAAATATRGCGAAAACTATCTTT
WI-5248	38	G	0.38	C	0.63	0.47	GCTGCTACGTTGTTASAGCAACCCCAGAAAA
WI-5248	99	C	0.31	T	0.69	0.43	TATTGACCGTACTTGYTCTTTGCTTTTTTTT
WI-5252	119	A	0.94	C	0.06	0.12	GTGAATCATTGCTTTMTACCATGTACATATT
WI-5257	77	C	0.75	A	0.25	0.38	CATGAAGCAAAGAGGMCTTTCATCTGCCCCT
WI-5300	38	T	0.88	C	0.13	0.22	GAGACCACTTCATTCYTTTTTGGATTATGAA
WI-5317	139	T	0.56	C	0.44	0.49	CTGGTAGCAGGTATAYGGACTCATTTCTTCT
WI-5328	44	A	0.94	G	0.06	0.12	ACACTGAAAAGACAGRAAAAAAAAAGAATATT
WI-5345	29	G	0.94	A	0.06	0.12	AGTTTTAAAAATCCTRCCTGCTATGGTTTGC
WI-5370	143	T	0.75	C	0.25	0.38	TAACTAATAAAACAAYTTTGAAATTCTCTGT
WI-5406	42	A	0.94	G	0.06	0.12	AGACTCTTCCAGAAGRGCCACTTCCACAGAT
WI-5406	118	C	0.63	A	0.38	0.47	TGTCAAGGTGAGAAAMCCTATGAGCCCACAC
WI-5406	120	C	0.81	T	0.19	0.30	TCAAGGTGAGAAACCYTATGAGCCCACACTT
WI-5415	54	T	0.75	A	0.25	0.38	TTCATCTTTCAGTTTWTAGATCGGATCATGA

WI-5437	41	C	0.19	T	0.81	0.30	AGAAAAATCCAAGAGYCTTAAACCATATTTT
WI-5481	29	G	0.44	A	0.56	0.49	TTAGTTGATGAATTTRAATTTTACAGTATCT
WI-5481	131	A	0.31	G	0.69	0.43	TTTATGCTGCAGTCGRAATACTTGGAGCCTG
WI-5492	38	T	0.94	C	0.06	0.12	CTTGTTAAAGTCCCAYCAAAGAAAGGATCCC
WI-5546	40	C	0.81	T	0.19	0.30	TGAAAAAAGGGAAAAYACCCATGTTTGCTAA
WI-5552	97	C	0.69	T	0.31	0.43	CAGCCTTTTTAGAGTYCCTGGGCAATTTGTG
WI-5573	58	C	0.75	T	0.25	0.38	ATAAGGAGGTGGGGAYGACACATTACTCTCC
WI-5612	44	T	0.94	A	0.06	0.12	TAAATCATTCTAACAWCACAAATATCTTATT
WI-5612	125	A	0.81	T	0.19	0.30	AGCATCGTGTCATTCWCAGTGTTTTAGGTTT
WI-5636	26	A	0.25	C	0.75	0.38	TTTATCCGCAATAAAMTTCCCAAAGTCCTCG
WI-5752	36	A	0.88	T	0.13	0.22	CTCAGTTTTTCCATCWTTTTTTCATAATTTA
WI-5791	44	C	0.94	G	0.06	0.12	TATTTGGATAAGTTTSACAAAGATGAGAACA
WI-5791	76	G	0.88	A	0.13	0.22	GTCCTAGAACCTCAGRATCGAAAGGAAGTTC
WI-5798	48	G	0.88	C	0.13	0.22	CCTTGTTTTCTTTTGSATTGAAAAATACTGG
WI-5836	161	C	0.94	T	0.06	0.12	ACATGATTCAATGATYCCATTTTGAAAATTA
WI-5850	92	C	0.94	T	0.06	0.12	GGCTTCCTCTATGCAYGCGTCTATCTTCTAT
WI-5850	134	G	0.88	A	0.13	0.22	TCCAATGTCCCATTCRTTTTGCCATTTCCTG
WI-5874	76	T	0.63	G	0.38	0.47	TACAGAAAAAAAAATTKTACATATCAAATGAC
WI-5944	52	A	0.69	G	0.31	0.43	ACCATGGGAATCTTGRTGCAAGTTAGATCCC
WI-5989	29	G	0.44	A	0.56	0.49	CAAAGGTCACAGGCARCGTACATACGGTTCT
WI-6053	24	A	0.94	G	0.06	0.12	GTGTCTAAGAACAACRTCTTCATGTCCAACT
WI-6141	80	T	0.88	C	0.13	0.22	TCTACAAGGTACTTAYCACTGTTCTGGGGTT
WI-6192	91	A	0.50	G	0.50	0.50	GGATTTAATTTGGATRATTTTAATACTTAGC
WI-6194	105	T	0.88	A	0.13	0.22	ATGATAATAAAGAAAWATGCAGACTACACTC
WI-6217	131	C	0.94	T	0.06	0.12	AGCAGCTCATTCAAGYGGCCCACCATGGCCC
WI-6272	86	C	0.31	T	0.69	0.43	AGGGAAAACTTTAATYTTCTTTGTCTTCTCC
WI-6303	96	G	0.63	A	0.38	0.47	AGAAGCTCTGTCTGCRCTGCAAAGCCATGGC
WI-6375	28	A	0.88	G	0.13	0.22	TATGGAAATCAATAGRTATCTTTTACAAAAA
WI-6409	73	A	0.94	T	0.06	0.12	CAAATCAATTACAACWATGTGCTTATCAGCT
WI-6409	112	T	0.69	A	0.31	0.43	ACCCCTATATTTTAAWGCAACTGACAGTTTT

WI-6450	45	T		0.63	G	0.38	0.47	CTATATCTTGTCACAKAGAAGTACCACACAT
WI-6461	88	C		0.94	T	0.06	0.12	TTCTATAAAACAACAYAAGGAACGAGGCTCA
WI-6523	165	G		0.69	T	0.31	0.43	TAGAGACTGAAGCTGKTATCAACCTTCCCTA
WI-6558	42	G		0.94	C	0.06	0.12	TTTATTAAGGACATTSTGTAATGTTTCCACT
WI-6558	68	C		0.56	T	0.44	0.49	CCACTTTGTTTTAAAYAATTACAAACATGTG
WI-6629	75	T		0.81	C	0.19	0.30	ATAAAAGTTGTCATAYAGCAATGGATGCTGT
WI-6686	151	A		0.44	G	0.56	0.49	CCAAAAACAAAGAATRAACATTGGAATAGTC
WI-6690	28	T		0.38	C	0.63	0.47	CATTATTAAGGAGAGYACTAGGAAAAACTAC
WI-6690	106	C		0.38	T	0.63	0.47	CTCTGGAGCCACAGCYGGCTAATACACTGCA
WI-6761	32	C		0.38	A	0.63	0.47	ACAGCTGCAGAATGGMCTTCTTCCTTCCCAG
WI-6770	53	A		0.13	G	0.88	0.22	CCCCAAAACATCACARAATTATTCATACTAT
WI-6889	139	T		0.88	C	0.13	0.22	ATGCAGTTAAAATTCYAGAATAATTAAAAGC
WI-7059	43	C		0.88	G	0.13	0.22	AGGCACCCAGCCATCSTGACCCAGCGAGGAG
WI-7254	37	A		0.75	G	0.25	0.38	TGAGAGAGGAGCCACRGTCCCTAATGACACC
WI-7286	65	T		0.44	C	0.56	0.49	AGCTTAACTGACAGAYGTTAAAGCTTTCTGG
WI-7374	182	T		0.94	A	0.06	0.12	TTGAAGAATATATTGWCAGAAACACAAGGCT
WI-7386	104	T		0.94	A	0.06	0.12	TGTAAACAATTGTTAWGTGTTTAGAATCAGA
WI-7423	107	T		0.44	C	0.56	0.49	GCTGGGCTGTGTTCCYCGGGCTCTTCTGGAC
WI-7424	131	T		0.44	A	0.56	0.49	GAGAGGAAAGAAAAAWACAACTTTCATTCTT
WI-7466	80	T		0.75	C	0.25	0.38	GGCTATGAAATAGTCYATTCAGTGAACTAGT
WI-7466	141	G		0.50	A	0.50	0.50	CAGTCTTTGTCCTGGRAATATCTCACAAAAT
WI-7593	46	G		0.06	A	0.94	0.12	AGGATGAAAGGAGAGRAATGAGATCAGTTTT
WI-7753	52	A		0.19	G	0.81	0.30	CCGAGAAGAACAGATRATCCCTGTATTTCAA
WI-7836	120	T		0.56	C	0.44	0.49	ACAATGCAACGTTCCYGATTTCTAATCTTGG
WI-7848	142	A		0.44	G	0.56	0.49	TTTTAAAACCGTCTCRTGTCTGAATAGCTTT
WI-7858	91	T		0.44	G	0.56	0.49	CGTGAATTTTTAAATKTATAGATGTAAACTT
WI-8172	136	C		0.63	G	0.38	0.47	TGTTTTCTTGACATASAGTACCTTTACAGGT
WI-8183	56	G		0.81	A	0.19	0.30	AACAATTTCTGTTGCRGCAGGTTTGATTTCA
WI-8377	63	A		0.94	G	0.06	0.12	CCCAGGCCCTTTCCCRTTATATCCAGGTATG
WI-8540	73	T		0.88	C	0.13	0.22	CCTGCATTTGGCTTAYGTGCCTGAAAAAGAA

WI-8550	32	G	0.50	A	0.50	0.50			TCAATGCAACAAGTARAATTTGTAAACTCAA
WI-8655	29	A	0.44	G	0.56	0.49			AATAGGAAACCAGAGRGGGAGCCCCAGGTGG
WI-8712	44	G	0.25	A	0.75	0.38			GAAGAGGTAGTGGAGRGAGATGGTCAGGCTT
WI-8827	22	C	0.19	T	0.81	0.30			CCTGGGAGACTATGGYAGTGAACACTAAAAT
WI-8833	51	A	0.88	G	0.13	0.22			CCATGCCATTCTCTGRTGCCCCTATAATGTG
WI-8850	21	A	0.50	G	0.50	0.50			CTTAACCTTTGGCCTRCCTGCCTGGCTGTTT
WI-8853	79	C	0.50	T	0.50	0.50			CGGGCATTGAGGATAYATGGAAGGCTCAGGA
WI-8865	42	T	0.31	C	0.69	0.43			TGAGGAAGACAGTCAYGGTCGAACAAACAAC
WI-8865	52	A	0.56	G	0.44	0.49			AGTCATGGTCGAACARACAACATGCTTCGGA
WI-8895	32	A	0.94	C	0.06	0.12			ACCAACCAACAGAATMCTCCCGTCCTTTGAA
WI-8974	34	C	0.38	T	0.63	0.47			GCCCTCAAGAACTCAYGCCAGCTCAGCCCTA
WI-8997	41	G	0.81	A	0.19	0.30			GCCCACTTGCTCCCCRTGAGCACTGCGTACA
WI-9005	26	C	0.81	T	0.19	0.30			TTTGCTGGGGAATCTYGTTTTTCTTCTTAAG
WI-9014	18	C	0.88	T	0.13	0.22			TGTTCCCATGCTGACYTGTGTTTCCTCCCCA
WI-9014	44	C	0.31	T	0.69	0.43			CCCCAGTCATCTTTCYTGTTCCAGAGAGGTG
WI-9014	93	T	0.63	C	0.38	0.47			TCTGTCTCAACTTTAYGTGCACTGAGCTGCA
WI-9015	48	C	0.00	T	1.00	0.00			AATTGGGCTGGATTGYGCTTTGGTTAATACA
WI-9063	53	A	0.44	C	0.56	0.49			AAAGACACCATTTATMTACCCAAGGGCAGAA
WI-9064	29	A	0.44	G	0.56	0.49			AAACATAATTGATTCRTATCTGCGAGACTTA
WI-9074	38	A	0.63	G	0.38	0.47			TTTGCTCTAAAAGAARAAGGAACTAGGTCAA
WI-9161	61	C	0.50	T	0.50	0.50			TAAGCATTGCCTGGCYTTCCTGTCTAGTCTC
WI-9171	62	G	0.94	A	0.06	0.12			TAGAGATAATAATCARTTCTTTACAACCGAT
WI-9174	47	T	0.56	C	0.44	0.49			CCATTCTCCTATTTAYCAGTCCTGTCCTATA
WI-9186	76	G	0.63	A	0.38	0.47			CCACTTCTCCCCGCARACCTAGGTCAGACTT
WI-9193	94	G	0.69	A	0.31	0.43			GTCTGCCTTAAAGCARTACCCCCCTACCACA
WI-9231	32	G	0.75	C	0.25	0.38			GGTCCCCCAGATTGASGTCTGAGTGTGGGCA
WI-9274	25	C	0.44	T	0.56	0.49			GACTTCACTTTGGTGYCAATGGACAGAAAAT
WI-9281	68	G	0.94	A	0.06	0.12			CTTGCTGGCTACTGGRTGTTAGTTTGCAGTC
WI-9304	70	G	0.25	A	0.75	0.38			ATGATCACCGACTGARAATATTGTTTTACAA
WI-9343	78	C	0.81	T	0.19	0.30			CAACATCCTCTGCCAYACACAACAAAACGTA

WI-9357	75	A	0.94	G	0.06	0.12 GTTATTATGCTCTTARTGATTTACAGACTGA
WI-9360	79	T	0.69	C	0.31	0.43 TCTGCTTTAACTTGGYATTCCTCTAATTGTG
WI-9413	112	G	0.38	C	0.63	0.47 CTGCTATTCCCAGATSAAGATTTGGTGGAAG
WI-9557	74	C	0.88	T	0.13	0.22 GCCCAGCTACAGCCTYGGTGCATCTTAACCC
WI-9720	47	A	0.00	G	1.00	0.00 AAAATACCCTTCTCTRATAATTTAAGTAACC
WI-9720	55	A	0.00	G	1.00	0.00 CTTCTCTAATAATTTRAGTAACCAAAATATT
WI-10019	139	A	0.88	T	0.13	0.22 TATGTAGCAAATCTAWTCCCCTAAGCACAGT
WI-10020	39	T	0.56	C	0.44	0.49 GTATTAAATAAATTAYGTTAACTGGCTCTGA
WI-10020	122	T	0.88	A	0.13	0.22 AAATCATGACTTTTTWAAAAATACCAGACTA
WI-10064	54	C	0.81	A	0.19	0.30 CAGGATCAGGGAAGGMATTATAATAAATATA
WI-10064	170	C	0.81	T	0.19	0.30 TGATTGTTTTACATGYGAAATCTGGCTTCAG
WI-10289	29	T	0.31	C	0.69	0.43 GTCCCCAAACTCTTAYTTAATTCCATTCAAT
WI-10316	104	T	0.44	C	0.56	0.49 ACCTCTATTCTCTTAYTAAACTTTTGGATAC
WI-10368	31	C	0.50	T	0.50	0.50 CAACCAGGTCTTGTTYCTACCCCTCTTAGAG
WI-10391	32	A	0.88	G	0.13	0.22 CAGGTATGACTCCCARTCAACTTCTTGACTC
WI-9748	74	C	0.94	G	0.06	0.12 TTACCCTTTGTCATTSTCAGACCAAGTACAT
WI-9763	21	G	0.75	A	0.25	0.38 AAACTCTGCGGTGTGRAGAAAGGACAGTTAT
WI-9897	83	A	0.63	T	0.38	0.47 ATTTATCTAGCCTGTWCAAGTCATCCAGTGA
WI-9897	84	C	0.88	T	0.13	0.22 TTTATCTAGCCTGTAYAAGTCATCCAGTGAG
WI-9935	42	C	0.56	T	0.44	0.49 TAATAACGTGTTGCAYACCTCACCAGAACTG
WI-9935	115	C	0.38	A	0.63	0.47 GGGGGAGTTCAGACAMAGCCAAGAAAAGCCT
WI-9943	91	T	0.81	C	0.19	0.30 TTTATATCCATCTTCYATTTTAATTTTCTAC
WI-10567	60	T	0.13	C	0.88	0.22 AAATATTATTCTTTTYTCATATTTTCCAATT
WI-10567	82	A	0.94	C	0.06	0.12 TTTCCAATTATTAATMCTAGAATTTTCACCA
WI-10567	146	A	0.13	C	0.88	0.22 GTCTTCTAATAGCAAMAGCTACTGGAAGCGG
WI-10686	133	C	0.81	T	0.19	0.30 TGCCCCTGTCCAAGGYTGTGTCTACACATGA
WI-10694	144	A	0.75	G	0.25	0.38 GCTTTATGAGTTTTCRTTTCCTCCTTTACAA
WI-10719	115	T	0.56	C	0.44	0.49 TCAAGGCCATTCTAGYGGCTGCTGGCAGTGC
WI-10721	40	A	0.38	G	0.63	0.47 CTCTGCTACTTGCCARATGAGATTTATTTAT
WI-10732	80	C	0.63	A	0.38	0.47 CTTCATTGGTTCACTMTTAAAGTTCTGTTAT

SUBSTITUTE SHEET (RULE 26)

WI-10775	39	C	0.75	T	0.25	0.38	TAATTCATTACACTCYACATCATATTTTCTT
WI-10778	62	A	0.13	G	0.88	0.22	GAGGAACATTTACAGRGTCCATCTCTGATGT
WI-10789	21	C	0.50	T	0.50	0.50	ACACTGCTCTAGACCYTCCCAGGGTCCCTCA
WI-10810	58	C	0.50	T	0.50	0.50	TCATGGGCAGGAATTYCATTTCTGTGTTTCT
WI-10828	23	T	0.94	C	0.06	0.12	CAGAATTACTTGGCAYAGGGTTTCTTAAAAC
WI-10832	91	G	0.75	C	0.25	0.38	ATCTGCAGGCTCTCCSTTTCTAAGTCACCTG
WI-10834	96	C	0.44	T	0.56	0.49	CAAAAGTGTGTTAATYCTTAATACCAATTTT
WI-11027	90	T	0.44	A	0.56	0.49	TACGCTTTTAAAAAAWAATAAAAATACTGTA
WI-11049	95	C	0.06	T	0.94	0.12	TGTTTCAACTAAGGAYAGACTTCAGAAGGCA
WI-11070	110	G	0.38	T	0.63	0.47	TCAGCCAGCTATCTTKGGTGCAGAGAGGTAC
WI-11070	135	C	0.75	T	0.25	0.38	AGGTACTCCAAGTACYGTGGGGGTTCTGATG
WI-11076	106	T	0.50	C	0.50	0.50	AAGGGGGAGCAGGCAYGTCACATACCCAGAG
WI-11076	142	G	0.81	A	0.19	0.30	GAGAGAGAAAGAGAGRAAGTGCCACACATTT
WI-11153	33	C	0.69	A	0.31	0.43	CTCACCTAAATTATGMGTGATTAAAATATAC
WI-11153	84	C	0.69	G	0.31	0.43	GCTTTAAGTACTTTASGAAGACCTTGACTGT
WI-11163	58	C	0.56	T	0.44	0.49	ATGACCAAAATGAGAYAAATTTGTTAAAAAA
WI-11169	95	A	0.75	G	0.25	0.38	AAAAAATTTAAGCCTRAAGTAGTGCTTTTTA
WI-11169	154	T	0.81	G	0.19	0.30	AAAAAAGAGCAGACAKTTTATCATGTGTTCT
WI-11175	77	T	0.81	A	0.19	0.30	TTTCTGCTCAAAGAGWTTTTTTTAAGTTATC
WI-11204	80	T	0.69	A	0.31	0.43	TGAAAAGAAAAACTTWCACCTTTTATTTTAA
WI-11204	88	T	0.94	C	0.06	0.12	AAAACTTTCACCTTTYATTTTAAAGTAACAT
WI-11206	127	A	0.81	T	0.19	0.30	CTGTATGTACAACTCWCCAACCATTAGGATT
WI-11215	68	C	0.94	T	0.06	0.12	CAGATTTATTTTAGTYATTTTTTTCTATAAT
WI-11219	18	G	0.56	A	0.44	0.49	AAAAATGCATTAGAARAATTGGAGGATAAAA
WI-11219	89	G	0.81	A	0.19	0.30	AGATGAAAAATAGGARAGAAAGTGTAGAAAA
WI-11222	25	C	0.75	T	0.25	0.38	GAATCATTTACACTAYCGAAATCAGCAAATG
WI-11222	136	G	0.88	A	0.13	0.22	TACCACTGCGGCTGGRTCACAACTTGGCTAC
WI-11226	165	A	0.94	C	0.06	0.12	TTTGGACTATGAACAMGACATAGTTGCTAAG
WI-11276	41	A	0.44	G	0.56	0.49	CAGCCAGGAGCAGACRCACCGGCTCCTCAGT
WI-11282	42	C	0.81	G	0.19	0.30	CAGAGAGCAAGGGAASCACACAAAATTTACA

WI-11295	37	A		0.56	G		0.44	0.49	AAAATATAATTTGCTRTAGAGTTCACAGATG
WI-11305	87	C		0.81	T		0.19	0.30	CACAGCATCACACCAYAGGGCCCACGGGAGG
WI-11321	67	A		0.56	G		0.44	0.49	AATAAATTTTTTTAARAAGGTTTAGCTATTC
WI-11324	40	C		0.56	G		0.44	0.49	AAATCATGTGCCCCASAGAGCCCCAAAGCTT
WI-11352	69	T		0.75	C		0.25	0.38	GCACATAGTGGAAAGYGCTAAGTGTCCTACG
WI-11352	104	T		0.75	C		0.25	0.38	GTCAGATCATATCCAYAGAAAAAACAGCTCTC
WI-11371	84	C		0.56	T		0.44	0.49	GAGATTCTGATTCAGYGTGCTCAGGCGGGGC
WI-11385	75	T		0.44	C		0.56	0.49	TAAAAGTCTCTTCAGYAGGAAAAAAGCTACA
WI-11388	88	C		0.25	A		0.75	0.38	CACGTAACTAAGTTCMTATAATTTTAACTTG
WI-11392	55	T		0.31	G		0.69	0.43	AACTTTAATAAATACKCTTTTTACAAAACAC
WI-11396	52	A		0.50	T		0.50	0.50	TTGAAATGGTGTTTTWGATGGGTGAATATGA
WI-11441	100	C		0.50	A		0.50	0.50	TCCCCACCAACCAGCMCAAATAAGGCCCTGG
WI-11466	26	C		0.69	T		0.31	0.43	CCATTTATTTTGCAGYCTTCAGTCCAAAAAA
WI-11537	119	C		0.88	G		0.13	0.22	TCTTACTCTGACCATSATAATCATTCTTTTT
WI-11549	102	T		0.44	G		0.56	0.49	TCTTTTAAATATCTGKGGGGATTTGTACAGA
WI-11585	79	T		0.63	C		0.38	0.47	TTTGCAAAAACAAAAYGGAAGTATCAGTGAA
WI-11604	68	G		0.94	C		0.06	0.12	CAGTTACCAGCATTTSAGAACTAGGGACTTT
WI-11614	60	A		0.75	G		0.25	0.38	AGACTCAGCTGCTTGRGGCATGTTCCCACCC
WI-11614	108	C		0.75	A		0.25	0.38	ACTGTGAAACTGCAAMATATTAAGTATTCGT
WI-11626	39	G		0.50	A		0.50	0.50	GGAACATGAAGGTAGRGATAAGTGTACAGGA
WI-11626	83	T		0.38	C		0.63	0.47	TATTTTTAAAATAAAYTACTTAATAATAAGA
WI-11627	23	T		0.69	C		0.31	0.43	CCTTCCATTGTCCTCYCTTGAGATGGGTTGC
WI-11636	61	A		0.88	G		0.13	0.22	AGATCTGCTTATCCTRTATATCCACATAACT
WI-11654	37	G		0.75	C		0.25	0.38	ACTATTCAGCAACTGSAAACTGTCCTGGGAG
WI-11656	28	G		0.25	A		0.75	0.38	TAGAAGGAACTGCAARCTTTACTTGAGGACA
WI-11680	55	T		0.94	C		0.06	0.12	TGATTCTCCCCTTTTYTTGCATAAAGGCTGG
WI-11696	47	T		0.88	C		0.13	0.22	CACAGCAGGGGACAGYAAGGTTGGCTTCTCT
WI-11702	69	C		0.81	T		0.19	0.30	AAATAACCACAGCAGYTTTCAGTATAATTTG
WI-11706	60	C		0.50	T		0.50	0.50	GTACAATTTATTTGCYGGCTGGAATTTGTTC
WI-11709	105	T		0.44	A		0.56	0.49	CTTGCTTCAGTTTGCWGTCCCGTAAAATTAG

WI-11710	103	C	0.50	A	0.50	0.50	AGCCTCAGTCTTCACMCTCCTCCCTCCTCCA
WI-11715	49	A	0.75	C	0.25	0.38	TGTAAAACAGACAAAMTGCATTACAACTGTG
WI-11715	123	C	0.63	T	0.38	0.47	GGCTGGCTGCAGCTTYAGCCACAGGATGGGG
WI-11727	43	G	0.38	C	0.63	0.47	AAACAACTATCAACASCTGCAACACAAACCA
WI-11728	16	C	0.50	G	0.50	0.50	TTTATTTATCAAACTSCAATTCCATTTCACA
WI-11758	61	A	0.88	G	0.13	0.22	TGTGGTTTTCGCCTGRTAGACCACAGGGCCA
WI-11773	93	T	0.06	C	0.94	0.12	CCTTTTTTTTCCCCCYGTGATTGTTAATTAG
WI-11790	28	A	0.81	G	0.19	0.30	TTACCAAACCTCTGTRGCTTAGCCTCGCCTA
WI-11806	60	T	0.88	G	0.13	0.22	AGAGTGGGCAGTTCAKGTTTTATTAGTATAT
WI-11879	61	C	0.81	A	0.19	0.30	GTATTTAGTATACAGMAGTGATTTTCTCTCT
WI-11906	52	A	0.69	G	0.31	0.43	AGAAAGAATCTGAATRTGAGGGAACTGCAGA
WI-11909	78	A	0.38	G	0.63	0.47	TGTTGGGTGGTCAAGRCTATTCAGAAAATCT
WI-11946	31	C	0.94	A	0.06	0.12	CTTTGTCCTGGAGACMCCAGCTAGTCTAAGA
WI-11965	65	T	0.56	G	0.44	0.49	CTCTGGTTTATTTAAKATCAACATTCACCAC
WI-12002	30	C	0.13	G	0.88	0.22	GAATCCAGGACACAASAAGAAAAACACCCAA
WI-12002	68	G	0.13	A	0.88	0.22	ATGGAGACAGAAGACRAGACACAACTCCTCC
WI-12002	89	T	0.56	C	0.44	0.49	CAACTCCTCCCCCACYGCCTCCCTGCTCTAG
WI-12018	31	A	0.56	T	0.44	0.49	AGCCAGCTCTGACTTWCTCTCTGTTTCTGTC
WI-12020	121	T	0.94	C	0.06	0.12	GAATACATGACCATTYCTCTTTTAGCACGTT
WI-12075	103	G	0.50	A	0.50	0.50	GGGCACGGGGGAGGCRGAAGGAAGAGAAAGA
WI-12086	72	C	0.81	T	0.19	0.30	GGAAAACTTGGATTTYCCAAGACCCGAAGAC
WI-12108	40	C	0.31	T	0.69	0.43	TTAAACTCAAATATCYGAAATACTTTCATTA
WI-12159	28	C	0.50	T	0.50	0.50	ACACCGTGCAAATGCYAAAGTGCACTGAGGA
WI-12169	121	G	0.81	C	0.19	0.30	TATTTTCTTTTGCTTSTTTTTTCTTTCACCT
WI-12173	57	C	0.88	T	0.13	0.22	TACAAAAAAATCCTGCYCTTATAGAGCATACA
WI-12179	96	G	0.50	A	0.50	0.50	GTACGGTGGAGGTCARGCATCTACAGGGTCA
WI-12201	61	C	0.69	T	0.31	0.43	CTGATCACCTGCATGYGCCAGGTATGTGGTC
WI-12210	76	A	0.88	G	0.13	0.22	AAACAACTATTGCATRGGAAAACATATGCAA
WI-12229	89	T	0.75	G	0.25	0.38	AAAAAGAGTAAAAATKACCAAAAAATTAAAG
WI-12234	66	A	0.44	G	0.56	0.49	ACACTTGTGGGGCTTRTTCAAACATGGACTG

WI-12310	46	G	0.88	A	0.13	0.22	TAATTTTAAAAAGCTRTTTAGGACCCAAACA
WI-12319	109	T	0.88	C	0.13	0.22	GTTCTGCTCATAATTYCCAATATGTACCAGA
WI-12323	68	G	0.50	A	0.50	0.50	GTACCTATGAAATAARACAGGTAGGGAATAT
WI-12326	25	G	0.81	A	0.19	0.30	TCAAAAGCAATTCACRCTTCCAGAATACAAA
WI-12340	18	T	0.94	C	0.06	0.12	CAATATAATTCCATTYCGAGTGATTAAAACC
WI-12345	37	C	0.50	A	0.50	0.50	CAGGAAAAAGAGGAAMCCTGAACCCCTCTGC
WI-12361	63	C	0.00	T	1.00	0.00	CAGCATATGTATTATYTGAACTAAATTTACA
WI-12469	91	C	0.56	T	0.44	0.49	TATATTCTATTTCTAYTTGACAGCACAGTTC
WI-12535	50	A	0.88	T	0.13	0.22	TTGAGGTGTAGATATWCTTCCTCTCTTCTCG
WI-12542	45	C	0.25	T	0.75	0.38	TGAACATTTAAATGTYATCCATGTGAGGGCT
WI-12542	70	G	0.50	T	0.50	0.50	AGGGCTCTAGATCATKGTAGGTGATTGATAC
WI-12542	71	G	0.63	T	0.38	0.47	GGGCTCTAGATCATGKTAGGTGATTGATACA
WI-12578	37	C	0.50	T	0.50	0.50	CTAAAGGAATGGGAAYGTGTTGGTGGTCGCT
WI-12601	42	T	0.56	C	0.44	0.49	TATTCTTGCTTTGATYGTCTACGTAAGCATG
WI-12634	52	T	0.31	C	0.69	0.43	TGTCTAGCAGTATTAYGCTATTAGCTATGTT
WI-12648	41	A	0.38	G	0.63	0.47	TGGCATTAAGGATGCRGTAGGATGTCCACTT
WI-12684	64	G	0.19	T	0.81	0.30	TGTAAACAGCTGTGCKCCATTTAGGCTTTGT
WI-12837	87	A	0.13	G	0.88	0.22	TCAAGGTAAAGTCCARTACAAAAAAACAGCA
WI-12988	36	C	0.56	A	0.44	0.49	GTGCTCTCAGTACAAMAAACAGCATCAGTAG
WI-13020	108	G	0.81	A	0.19	0.30	AACCCTGAGACTTTARATCTGCAAAGGGGTT
WI-13112	71	C	0.13	T	0.88	0.22	GACTTAAGCTTTTTTYCTTTTTCCATATAAT
WI-13119	51	C	0.94	G	0.06	0.12	GACACAATCAAGACTSACAGTAGCCTCAACC
WI-13119	114	G	0.88	C	0.13	0.22	GGACTACAGGCATGTSACACCACACCTGGTT
WI-13264	25	G	0.31	A	0.69	0.43	AAGGCTCTTGCCCATRTATTCCCGTCTCTCC
WI-13364	35	A	0.38	G	0.63	0.47	TTTTTTAGTAGAAGCRGGAACAGTTGTCAAT
WI-13367	84	C	0.44	G	0.56	0.49	GAAGACTCACCAGAASAGGGTGGGGTGGGGA
WI-13373	52	G	0.94	A	0.06	0.12	GAATAAACATCTCACRAACTGTCGCTCCTAG
WI-13416	71	C	0.50	A	0.50	0.50	TGACAAGAACACATAMAAATATTGAAATTAT
WI-13424	66	G	0.88	A	0.13	0.22	TTCACCCTATTCTTCRTAGACCCTGGGGAGA
WI-13446	22	G	0.50	C	0.50	0.50	TTCTTTCACTCATCASCCTTCTGATTTTGAT

SUBSTITUTE SHEET (RULE 26)

WI-13453	88	T	0.63	A	0.38	0.47	AAATCTTGTCTCTTCWTGCTAGAAAGAGATG
WI-13470	100	C	0.81	A	0.19	0.30	ATATTGGAATTTCTAMAGAGACCCATGGTCT
WI-13473	31	C	0.94	T	0.06	0.12	ATGGGCTGAGACTGTYTGTCTGGTAGATGCA
WI-13477	32	A	0.44	G	0.56	0.49	TTGTTGGATAAAAGGRCATTGTTTTTCATTA
WI-13477	61	A	0.88	G	0.13	0.22	TAGCTTGTCTTCAAARGACAGAGAAATAAGA
WI-13507	41	T	0.94	C	0.06	0.12	AGCTTGACCTTAGGTYAATATTTCATTTGGG
WI-13522	33	C	0.31	T	0.69	0.43	CCCCACTAATACAACYGAGAACCACTGACTT
WI-13528	80	A	0.44	G	0.56	0.49	AAAAAGAAGACATTTRTTCAGAGAAAACTGT
WI-13529	42	T	0.75	C	0.25	0.38	ATTGAACAGTTACCAYAAGCAAGAGAGTGAG
WI-13536	29	T	0.94	C	0.06	0.12	AAAAACTCAGCGAAGYGAAAAGGTGGATAGC
WI-13551	74	G	0.75	A	0.25	0.38	TATATTCAGACAATCRAATATTACTTAGCAC
WI-13578	48	T	0.63	A	0.38	0.47	AGCAGAAAGAAAAACCWAGACAAAAAGATGTT
WI-13582	43	C	0.88	A	0.13	0.22	TCTAGAGACTGGGGAMTGGAATCTAACTGCG
WI-13594	66	G	0.75	A	0.25	0.38	CAGATCACACAAAAAGCRTGCACACAAAAAAGTAC
WI-13600	26	G	0.88	T	0.13	0.22	GAGCCAAGCATCCATKCCATCATCTAGTAAC
WI-13602	89	G	0.75	T	0.25	0.38	TCTGGAGACAACACACAKAAATCTATTAATATT
WI-13650	76	A	0.56	T	0.44	0.49	TTTCACTTTTAAAACWTAAAAAACTACTCTT
WI-13654	49	A	0.63	G	0.38	0.47	TGAAACACATCCGTARGTATGACATCATTTC
WI-13683	47	C	0.94	G	0.06	0.12	ACCTATCTGCCCATGSTTTACAGCCTTTTAA
WI-13712	40	A	0.69	C	0.31	0.43	ATTTTTATTCTATTGMATTATAAGAAAAGTG
WI-13725	56	A	0.88	C	0.13	0.22	GCACATATGGGTGCCMGCCCGAGACAGCAGG
WI-13744	115	C	0.38	T	0.63	0.47	CTGAACAAAACTGAAYGCTGTGCTTATCTTT
WI-13752	106	T	0.81	C	0.19	0.30	AAGTGCTGGATATACYTGGCTTGCACCGGAC
WI-13752	117	C	0.31	T	0.69	0.43	ATACTTGGCTTGCACYGGACACCTTTTACGG
WI-13763	59	T	0.88	C	0.13	0.22	GGACACTGCAGTGATYAGGGGCAGGTGTGGG
WI-13785	27	T	0.31	C	0.69	0.43	ACTATAAAAGTGCTTYAAAATGCAGCAGCAG
WI-13785	40	C	0.38	G	0.63	0.47	TTTAAAATGCAGCAGSAGGAGATGTGAAGAC
WI-13785	56	A	0.56	C	0.44	0.49	AGGAGATGTGAAGACMCAAATGAACAAGTGC
WI-13785	72	G	0.56	A	0.44	0.49	CAAATGAACAAGTGCRTAGTGACACATAGCT
WI-13789	62	G	0.63	A	0.38	0.47	GGATGGCTGAGGGAGRGAACAGAGGAAGCGC

WI-13793	88	C	0.31	G	0.69	0.43	CAGCCTAGATATAGGSAGTAACAAATCCTCC
WI-13794	52	A	0.44	G	0.56	0.49	ACCCTTTTCTTTCTCRTACAAGGTTAAGAGC
WI-13805	112	G	0.44	A	0.56	0.49	AAGGCACACGGGGAARGGGTCAAGGCAGGCT
WI-13805	113	G	0.44	A	0.56	0.49	AGGCACACGGGGAAGRGGTCAAGGCAGGCTG
WI-13806	62	G	0.94	A	0.06	0.12	AACTAGGCCTCAGGTRCCCATTAAGCATGCT
WI-13810	106	T	0.81	C	0.19	0.30	ATACATCCAAAACTTYAGTTAGCAGCAAGCA
WI-13831	56	G	0.94	C	0.06	0.12	AGGTGACTTGGAAAASGAGATTCACATACTT
WI-13831	113	T	0.25	C	0.75	0.38	CTTCTCTTCTGTAGAYGTCTCCATGTTACAG
WI-13850	51	A	0.88	G	0.13	0.22	TTTTAACACAGCCATRTTACAAACATTGTCA
WI-13857	28	A	0.94	G	0.06	0.12	AATGCTTTTCTGAACRTACATTTTAGGTATC
WI-13859	84	G	0.94	A	0.06	0.12	TGAAAAGGAAACTATRACAAACAAGTATATA
WI-13892	50	G	0.81	A	0.19	0.30	TTTTTAAATAGAACARCTTTGATTTTTTAGTA
WI-13909	80	G	0.88	A	0.13	0.22	ACTCTCTTCAAACTCRAATATCTTTTTTCAGA
WI-13909	93	A	0.88	T	0.13	0.22	TCGAATATCTTTTTCWGAGATGTCTAGCTAG
WI-13910	63	C	0.38	T	0.63	0.47	ACGTCCTTTGTGCTAYGTGATAAGTGTGCTT
WI-13936	123	C	0.81	T	0.19	0.30	ATTCAATAGCCTATCYAACTCCATGTGGGAG
WI-13951	39	C	0.63	T	0.38	0.47	AAGTAATGAACAAAAYAGACCCCAGATCAGA
WI-13951	88	G	0.63	C	0.38	0.47	GTTAATTCTGGAGCASATTCAAGCAGCAAAT
WI-13960	39	A	0.81	C	0.19	0.30	TTAAATACTGATAGAMGATGCAAATTTGTCC
WI-13967	103	A	0.56	C	0.44	0.49	ACAAGGAAATAAAAAMCACTTTTAGGAGATG
WI-13983	52	G	0.75	A	0.25	0.38	CCACTCCTTAAACCTRCCACTGGGCTAAGAG
WI-14061	68	C	0.94	T	0.06	0.12	CCGTACATACCTTATYAACCATTTCATCCAC
WI-14065	29	T	0.50	C	0.50	0.50	AGGTCAGAGGCAATTYGAGATCCCAGATTCA
WI-14078	61	C	0.19	T	0.81	0.30	TTAGGAAGAGCAAGAYGCAGTAAGAGACATG
WI-14083	47	C	0.31	T	0.69	0.43	GCTTAAAACAACACTYATTTGTTATTTCACA
WI-14085	31	A	0.13	G	0.88	0.22	TGTAAGAAGAAAAACRTAACTAGCACGTGAA
WI-14102	22	C	0.50	A	0.50	0.50	AAACAAAGCAGAAAAMCCCACCATTAACAAG
WI-14124	92	A	0.94	G	0.06	0.12	CGTTAACACTAAGCCRTATTATTCAAAATGT
WI-14125	88	C	0.63	T	0.38	0.47	ATTTTTTGACGACTAYGTGGCCATGCCATTC
WI-14136	120	G	0.75	A	0.25	0.38	ACCATGTCTTCACATRGCCCAAAGAGACAGA

WI-14138	23	C	0.88	T	0.13	0.22	GGCACCAGAAAAGCTYATGTTCTATGTTATG	
WI-14149	83	C	0.94	T	0.06	0.12	TTAGCGTTAAAGGAGYTGAGTTGAGTCAAAC	
WI-14153	28	A	0.56	G	0.44	0.49	TGCAGGAAGGCCAGCRTCCCCTCCTGCCGTT	
WI-14162	57	A	0.81	G	0.19	0.30	TGGCCTCGCTGCCTCRGCCTTTTCTCTTTGA	
WI-14186	52	C	0.50	T	0.50	0.50	ATGGAAAGACACATAYGGTACAAAATTACAG	
WI-14186	88	A	0.50	G	0.50	0.50	TTAGTTCATTACATGRTACAAATCATTAGAG	

Analysis of Polymorphisms

A. Preparation of Samples

Polymorphisms are detected in a target nucleic acid from an individual being analyzed. For assay of genomic DNA, virtually any biological sample (other than pure red blood cells) is suitable. For example, convenient tissue samples include whole blood, semen, saliva, tears, urine, fecal material, sweat, buccal, skin and hair. For assay of cDNA or mRNA, the tissue sample must be obtained from an organ in which the target nucleic acid is expressed. For example, if the target nucleic acid is a cytochrome P450, the liver is a suitable source.

Many of the methods described below require amplification of DNA from target samples. This can be accomplished by e.g., PCR. *See generally PCR Technology: Principles and Applications for DNA Amplification* (ed. H.A. Erlich, Freeman Press, NY, NY, 1992); *PCR Protocols: A Guide to Methods and Applications* (eds. Innis, et al., Academic Press, San Diego, CA, 1990); Mattila et al., *Nucleic Acids Res.* 19, 4967 (1991); Eckert et al., *PCR Methods and Applications* 1, 17 (1991); *PCR* (eds. McPherson et al., IRL Press, Oxford); and U.S. Patent 4,683,202 (each of which is incorporated by reference for all purposes).

Other suitable amplification methods include the ligase chain reaction (LCR) (see Wu and Wallace, *Genomics* 4, 560 (1989), Landegren et al., *Science* 241, 1077 (1988), transcription amplification (Kwoh et al., *Proc. Natl. Acad. Sci. USA* 86, 1173 (1989)), and self-sustained sequence replication (Guatelli et al., *Proc. Nat. Acad. Sci. USA,* 87, 1874 (1990)) and nucleic acid based sequence amplification (NASBA). The latter two amplification methods involve isothermal reactions based on isothermal transcription, which produce both single stranded RNA (ssRNA) and double stranded DNA (dsDNA) as the amplification products in a ratio of about 30 or 100 to 1, respectively.

B. Detection of Polymorphisms in Target DNA

There are two distinct types of analysis depending whether a polymorphism in question has already been characterized. The first type of analysis

is sometimes referred to as de novo characterization. This analysis compares target sequences in different individuals to identify points of variation, i.e., polymorphic sites. By analyzing a groups of individuals representing the greatest ethnic diversity among humans and greatest breed and species variety in plants and animals, patterns characteristic of the most common alleles/haplotypes of the locus can be identified, and the frequencies of such populations in the population determined. Additional allelic frequencies can be determined for subpopulations characterized by criteria such as geography, race, or gender. The de novo identification of the polymorphisms of the invention is described in the Examples section. The second type of analysis is determining which form(s) of a characterized polymorphism are present in individuals under test. There are a variety of suitable procedures, which are discussed in turn.

1. Allele-Specific Probes

The design and use of allele-specific probes for analyzing polymorphisms is described by e.g., Saiki et al., *Nature* 324, 163-166 (1986); Dattagupta, EP 235,726, Saiki, WO 89/11548. Allele-specific probes can be designed that hybridize to a segment of target DNA from one individual but do not hybridize to the corresponding segment from another individual due to the presence of different polymorphic forms in the respective segments from the two individuals. Hybridization conditions should be sufficiently stringent that there is a significant difference in hybridization intensity between alleles, and preferably an essentially binary response, whereby a probe hybridizes to only one of the alleles. Some probes are designed to hybridize to a segment of target DNA such that the polymorphic site aligns with a central position (e.g., in a 15 mer at the 7 position; in a 16 mer, at either the 8 or 9 position) of the probe. This design of probe achieves good discrimination in hybridization between different allelic forms.

Allele-specific probes are often used in pairs, one member of a pair showing a perfect match to a reference form of a target sequence and the other member showing a perfect match to a variant form. Several pairs of probes can then be immobilized on the same support for simultaneous analysis of multiple polymorphisms within the same target sequence.

2. Tiling Arrays

The polymorphisms can also be identified by hybridization to nucleic acid arrays, some example of which are described by WO 95/11995 (incorporated by reference in its entirety for all purposes). One form of such arrays is described in the Examples section in connection with de novo identification of polymorphisms. The same array or a different array can be used for analysis of characterized polymorphisms. WO 95/11995 also describes subarrays that are optimized for detection of a variant forms of a precharacterized polymorphism. Such a subarray contains probes designed to be complementary to a second reference sequence, which is an allelic variant of the first reference sequence. The second group of probes is designed by the same principles as described in the Examples except that the probes exhibit complementarily to the second reference sequence. The inclusion of a second group (or further groups) can be particular useful for analyzing short subsequences of the primary reference sequence in which multiple mutations are expected to occur within a short distance commensurate with the length of the probes (*i.e.*, two or more mutations within 9 to 21 bases).

3. Allele-Specific Primers

An allele-specific primer hybridizes to a site on target DNA overlapping a polymorphism and only primes amplification of an allelic form to which the primer exhibits perfect complementarily. See Gibbs, *Nucleic Acid Res.* 17, 2427-2448 (1989). This primer is used in conjunction with a second primer which hybridizes at a distal site. Amplification proceeds from the two primers leading to a detectable product signifying the particular allelic form is present. A control is usually performed with a second pair of primers, one of which shows a single base mismatch at the polymorphic site and the other of which exhibits perfect complementarily to a distal site. The single-base mismatch prevents amplification and no detectable product is formed. The method works best when the mismatch is included in the 3'-most position of the oligonucleotide aligned with the polymorphism because this position is most destabilizing to elongation from the primer. *See, e.g.*, WO 93/22456.

4. Direct-Sequencing

The direct analysis of the sequence of polymorphisms of the present invention can be accomplished using either the dideoxy chain termination method or the Maxam Gilbert method (see Sambrook et al., *Molecular Cloning, A Laboratory Manual* (2nd Ed., CSHP, New York 1989); Zyskind et al., *Recombinant DNA Laboratory Manual*, (Acad. Press, 1988)).

5. Denaturing Gradient Gel Electrophoresis

Amplification products generated using the polymerase chain reaction can be analyzed by the use of denaturing gradient gel electrophoresis. Different alleles can be identified based on the different sequence-dependent melting properties and electrophoretic migration of DNA in solution. Erlich, ed., *PCR Technology, Principles and Applications for DNA Amplification*, (W.H. Freeman and Co, New York, 1992), Chapter 7.

6. Single-Strand Conformation Polymorphism Analysis

Alleles of target sequences can be differentiated using single-strand conformation polymorphism analysis, which identifies base differences by alteration in electrophoretic migration of single stranded PCR products, as described in Orita et al., *Proc. Nat. Acad. Sci.* 86, 2766-2770 (1989). Amplified PCR products can be generated as described above, and heated or otherwise denatured, to form single stranded amplification products. Single-stranded nucleic acids may refold or form secondary structures which are partially dependent on the base sequence. The different electrophoretic mobilities of single-stranded amplification products can be related to base-sequence difference between alleles of target sequences.

III. Methods of Use

After determining polymorphic form(s) present in an individual at one or more polymorphic sites, this information can be used in a number of methods.

A. Forensics

Determination of which polymorphic forms occupy a set of polymorphic sites in an individual identifies a set of polymorphic forms that distinguishes the individual. *See generally* National Research Council, *The Evaluation of Forensic DNA Evidence* (Eds. Pollard et al., National Academy Press, DC, 1996). The more sites that are analyzed the lower the probability that the set of polymorphic forms in one individual is the same as that in an unrelated individual. Preferably, if multiple sites are analyzed, the sites are unlinked. Thus, polymorphisms of the invention are often used in conjunction with polymorphisms in distal genes. Preferred polymorphisms for use in forensics are diallelic because the population frequencies of two polymorphic forms can usually be determined with greater accuracy than those of multiple polymorphic forms at multi-allelic loci.

The capacity to identify a distinguishing or unique set of forensic markers in an individual is useful for forensic analysis. For example, one can determine whether a blood sample from a suspect matches a blood or other tissue sample from a crime scene by determining whether the set of polymorphic forms occupying selected polymorphic sites is the same in the suspect and the sample. If the set of polymorphic markers does not match between a suspect and a sample, it can be concluded (barring experimental error) that the suspect was not the source of the sample. If the set of markers does match, one can conclude that the DNA from the suspect is consistent with that found at the crime scene. If frequencies of the polymorphic forms at the loci tested have been determined (e.g., by analysis of a suitable population of individuals), one can perform a statistical analysis to determine the probability that a match of suspect and crime scene sample would occur by chance.

p(ID) is the probability that two random individuals have the same polymorphic or allelic form at a given polymorphic site. In diallelic loci, four genotypes are possible: AA, AB, BA, and BB. If alleles A and B occur in a haploid genome of the organism with frequencies x and y, the probability of each genotype in a diploid organism are (see WO 95/12607):

Homozygote: $p(AA) = x^2$

Homozygote: $p(BB) = y^2 = (1-x)^2$
Single Heterozygote: $p(AB) = p(BA) = xy = x(1-x)$
Both Heterozygotes: $p(AB+BA) = 2xy = 2x(1-x)$

The probability of identity at one locus (i.e, the probability that two individuals, picked at random from a population will have identical polymorphic forms at a given locus) is given by the equation:

$p(ID) = (x^2)^2 + (2xy)^2 + (y^2)^2$.

These calculations can be extended for any number of polymorphic forms at a given locus. For example, the probability of identity p(ID) for a 3-allele system where the alleles have the frequencies in the population of x, y and z, respectively, is equal to the sum of the squares of the genotype frequencies:

$p(ID) = x^4 + (2xy)^2 + (2yz)^2 + (2xz)^2 + z^4 + y^4$

In a locus of n alleles, the appropriate binomial expansion is used to calculate p(ID) and p(exc).

The cumulative probability of identity (cum p(ID)) for each of multiple unlinked loci is determined by multiplying the probabilities provided by each locus.

cum p(ID) = p(ID1)p(ID2)p(ID3).... p(IDn)

The cumulative probability of non-identity for n loci (i.e. the probability that two random individuals will be different at 1 or more loci) is given by the equation:

cum p(nonID) = 1-cum p(ID).

If several polymorphic loci are tested, the cumulative probability of non-identity for random individuals becomes very high (e.g., one billion to one). Such probabilities can be taken into account together with other evidence in determining the guilt or innocence of the suspect.

B. Paternity Testing

The object of paternity testing is usually to determine whether a male is the father of a child. In most cases, the mother of the child is known and thus, the mother's contribution to the child's genotype can be traced. Paternity testing

investigates whether the part of the child's genotype not attributable to the mother is consistent with that of the putative father. Paternity testing can be performed by analyzing sets of polymorphisms in the putative father and the child.

If the set of polymorphisms in the child attributable to the father does not match the putative father, it can be concluded, barring experimental error, that the putative father is not the real father. If the set of polymorphisms in the child attributable to the father does match the set of polymorphisms of the putative father, a statistical calculation can be performed to determine the probability of coincidental match.

The probability of parentage exclusion (representing the probability that a random male will have a polymorphic form at a given polymorphic site that makes him incompatible as the father) is given by the equation (see WO 95/12607):

$$p(exc) = xy(1-xy)$$

where x and y are the population frequencies of alleles A and B of a diallelic polymorphic site.

(At a triallelic site $p(exc) = xy(1-xy) + yz(1- yz) + xz(1-xz)+ 3xyz(1-xyz)))$, where x, y and z and the respective population frequencies of alleles A, B and C).

The probability of non-exclusion is

$$p(non\text{-}exc) = 1\text{-}p(exc)$$

The cumulative probability of non-exclusion (representing the value obtained when n loci are used) is thus:

$$cum\ p(non\text{-}exc) = p(non\text{-}exc1)p(non\text{-}exc2)p(non\text{-}exc3)\ldots. p(non\text{-}excn)$$

The cumulative probability of exclusion for n loci (representing the probability that a random male will be excluded)

$$cum\ p(exc) = 1 - cum\ p(non\text{-}exc).$$

If several polymorphic loci are included in the analysis, the cumulative probability of exclusion of a random male is very high. This probability can be taken into account in assessing the liability of a putative father whose polymorphic marker set matches the child's polymorphic marker set attributable to his/her father.

C. Correlation of Polymorphisms with Phenotypic Traits

The polymorphisms of the invention may contribute to the phenotype of an organism in different ways. Some polymorphisms occur within a protein coding sequence and contribute to phenotype by affecting protein structure. The effect may be neutral, beneficial or detrimental, or both beneficial and detrimental, depending on the circumstances. For example, a heterozygous sickle cell mutation confers resistance to malaria, but a homozygous sickle cell mutation is usually lethal. Other polymorphisms occur in noncoding regions but may exert phenotypic effects indirectly via influence on replication, transcription, and translation. A single polymorphism may affect more than one phenotypic trait. Likewise, a single phenotypic trait may be affected by polymorphisms in different genes. Further, some polymorphisms predispose an individual to a distinct mutation that is causally related to a certain phenotype.

Phenotypic traits include diseases that have known but hitherto unmapped genetic components (e.g., agammaglobulimenia, diabetes insipidus, Lesch-Nyhan syndrome, muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial hypercholesterolemia, polycystic kidney disease, hereditary spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary hemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos syndrome, osteogenesis imperfecta, and acute intermittent porphyria). Phenotypic traits also include symptoms of, or susceptibility to, multifactorial diseases of which a component is or may be genetic, such as autoimmune diseases, inflammation, cancer, diseases of the nervous system, and infection by pathogenic microorganisms. Some examples of autoimmune diseases include rheumatoid arthritis, multiple sclerosis, diabetes (insulin-dependent and non-independent), systemic lupus erythematosus and Graves disease. Some examples of cancers include cancers of the bladder, brain, breast, colon, esophagus, kidney, leukemia, liver, lung, oral cavity, ovary, pancreas, prostate, skin, stomach and uterus. Phenotypic traits also include characteristics such as longevity, appearance (e.g., baldness, obesity), strength, speed, endurance, fertility, and susceptibility or receptivity to particular drugs or therapeutic treatments.

Correlation is performed for a population of individuals who have been

tested for the presence or absence of a phenotypic trait of interest and for polymorphic markers sets. To perform such analysis, the presence or absence of a set of polymorphisms (i.e. a polymorphic set) is determined for a set of the individuals, some of whom exhibit a particular trait, and some of which exhibit lack of the trait. The alleles of each polymorphism of the set are then reviewed to determine whether the presence or absence of a particular allele is associated with the trait of interest. Correlation can be performed by standard statistical methods such as a κ-squared test and statistically significant correlations between polymorphic form(s) and phenotypic characteristics are noted. For example, it might be found that the presence of allele A1 at polymorphism A correlates with heart disease. As a further example, it might be found that the combined presence of allele A1 at polymorphism A and allele B1 at polymorphism B correlates with increased milk production of a farm animal.

Such correlations can be exploited in several ways. In the case of a strong correlation between a set of one or more polymorphic forms and a disease for which treatment is available, detection of the polymorphic form set in a human or animal patient may justify immediate administration of treatment, or at least the institution of regular monitoring of the patient. Detection of a polymorphic form correlated with serious disease in a couple contemplating a family may also be valuable to the couple in their reproductive decisions. For example, the female partner might elect to undergo in vitro fertilization to avoid the possibility of transmitting such a polymorphism from her husband to her offspring. In the case of a weaker, but still statistically significant correlation between a polymorphic set and human disease, immediate therapeutic intervention or monitoring may not be justified. Nevertheless, the patient can be motivated to begin simple life-style changes (e.g., diet, exercise) that can be accomplished at little cost to the patient but confer potential benefits in reducing the risk of conditions to which the patient may have increased susceptibility by virtue of variant alleles. Identification of a polymorphic set in a patient correlated with enhanced receptiveness to one of several treatment regimes for a disease indicates that this treatment regime should be followed.

For animals and plants, correlations between characteristics and phenotype are useful for breeding for desired characteristics. For example, Beitz et

al., US 5,292,639 discuss use of bovine mitochondrial polymorphisms in a breeding program to improve milk production in cows. To evaluate the effect of mtDNA D-loop sequence polymorphism on milk production, each cow was assigned a value of 1 if variant or 0 if wildtype with respect to a prototypical mitochondrial DNA sequence at each of 17 locations considered. Each production trait was analyzed individually with the following animal model:

$$Y_{ijkpn} = \mu + YS_i + P_j + X_k + \beta_1 + \ldots \beta_{17} + PE_n + a_n + e_p$$

where Y_{ijknp} is the milk, fat, fat percentage, SNF, SNF percentage, energy concentration, or lactation energy record; μ is an overall mean; YS_i is the effect common to all cows calving in year-season; X_k is the effect common to cows in either the high or average selection line; β_1 to β_{17} are the binomial regressions of production record on mtDNA D-loop sequence polymorphisms; PE_n is permanent environmental effect common to all records of cow n; a_n is effect of animal n and is composed of the additive genetic contribution of sire and dam breeding values and a Mendelian sampling effect; and e_p is a random residual. It was found that eleven of seventeen polymorphisms tested influenced at least one production trait. Bovines having the best polymorphic forms for milk production at these eleven loci are used as parents for breeding the next generation of the herd.

D. Genetic Mapping of Phenotypic Traits

The previous section concerns identifying correlations between phenotypic traits and polymorphisms that directly or indirectly contribute to those traits. The present section describes identification of a physical linkage between a genetic locus associated with a trait of interest and polymorphic markers that are not associated with the trait, but are in physical proximity with the genetic locus responsible for the trait and co-segregate with it. Such analysis is useful for mapping a genetic locus associated with a phenotypic trait to a chromosomal position, and thereby cloning gene(s) responsible for the trait. *See* Lander et al., *Proc. Natl. Acad. Sci. (USA)* 83, 7353-7357 (1986); Lander et al., *Proc. Natl. Acad. Sci. (USA)* 84, 2363-2367 (1987); Donis-Keller et al., *Cell* 51, 319-337 (1987); Lander et al.,

Genetics 121, 185-199 (1989)). Genes localized by linkage can be cloned by a process known as directional cloning. *See* Wainwright, *Med. J. Australia* 159, 170-174 (1993); Collins, *Nature Genetics* 1, 3-6 (1992) (each of which is incorporated by reference in its entirety for all purposes).

Linkage studies are typically performed on members of a family. Available members of the family are characterized for the presence or absence of a phenotypic trait and for a set of polymorphic markers. The distribution of polymorphic markers in an informative meiosis is then analyzed to determine which polymorphic markers co-segregate with a phenotypic trait. *See, e.g.*, Kerem et al., *Science* 245, 1073-1080 (1989); Monaco et al., *Nature* 316, 842 (1985); Yamoka et al., *Neurology* 40, 222-226 (1990); Rossiter et al., *FASEB Journal* 5, 21-27 (1991).

Linkage is analyzed by calculation of LOD (log of the odds) values. A lod value is the relative likelihood of obtaining observed segregation data for a marker and a genetic locus when the two are located at a recombination fraction θ, versus the situation in which the two are not linked, and thus segregating independently (Thompson & Thompson, *Genetics in Medicine* (5th ed, W.B. Saunders Company, Philadelphia, 1991); Strachan, "Mapping the human genome" in *The Human Genome* (BIOS Scientific Publishers Ltd, Oxford), Chapter 4). A series of likelihood ratios are calculated at various recombination fractions (θ), ranging from $\theta = 0.0$ (coincident loci) to $\theta = 0.50$ (unlinked). Thus, the likelihood at a given value of θ is: probability of data if loci linked at θ to probability of data if loci unlinked. The computed likelihoods are usually expressed as the $\log_{10}$ of this ratio (i.e., a lod score). For example, a lod score of 3 indicates 1000:1 odds against an apparent observed linkage being a coincidence. The use of logarithms allows data collected from different families to be combined by simple addition. Computer programs are available for the calculation of lod scores for differing values of θ (e.g., LIPED, MLINK (Lathrop, *Proc. Nat. Acad. Sci. (USA)* 81, 3443-3446 (1984)). For any particular lod score, a recombination fraction may be determined from mathematical tables. *See* Smith et al., *Mathematical tables for research workers in human genetics* (Churchill, London, 1961); Smith, *Ann. Hum. Genet.* 32, 127-150 (1968). The value of θ at which the lod

score is the highest is considered to be the best estimate of the recombination fraction.

Positive lod score values suggest that the two loci are linked, whereas negative values suggest that linkage is less likely (at that value of θ) than the possibility that the two loci are unlinked. By convention, a combined lod score of +3 or greater (equivalent to greater than 1000:1 odds in favor of linkage) is considered definitive evidence that two loci are linked. Similarly, by convention, a negative lod score of -2 or less is taken as definitive evidence against linkage of the two loci being compared. Negative linkage data are useful in excluding a chromosome or a segment thereof from consideration. The search focuses on the remaining non-excluded chromosomal locations.

IV. Modified Polypeptides and Gene Sequences

The invention further provides variant forms of nucleic acids and corresponding proteins. The nucleic acids comprise one of the sequences described in Table 1, column 8, in which the polymorphic position is occupied by one of the alternative bases for that position. Some nucleic acid encode full-length variant forms of proteins. Similarly, variant proteins have the prototypical amino acid sequences of encoded by nucleic acid sequence shown in Table 1, column 8, (read so as to be in-frame with the full-length coding sequence of which it is a component) except at an amino acid encoded by a codon including one of the polymorphic positions shown in the Table. That position is occupied by the amino acid coded by the corresponding codon in any of the alternative forms shown in the Table.

Variant genes can be expressed in an expression vector in which a variant gene is operably linked to a native or other promoter. Usually, the promoter is a eukaryotic promoter for expression in a mammalian cell. The transcription regulation sequences typically include a heterologous promoter and optionally an enhancer which is recognized by the host. The selection of an appropriate promoter, for example trp, lac, phage promoters, glycolytic enzyme promoters and tRNA promoters, depends on the host selected. Commercially available expression vectors can be used. Vectors can include host-recognized replication systems, amplifiable genes, selectable markers, host

sequences useful for insertion into the host genome, and the like.

The means of introducing the expression construct into a host cell varies depending upon the particular construction and the target host. Suitable means include fusion, conjugation, transfection, transduction, electroporation or injection, as described in Sambrook, *supra*. A wide variety of host cells can be employed for expression of the variant gene, both prokaryotic and eukaryotic. Suitable host cells include bacteria such as *E. coli*, yeast, filamentous fungi, insect cells, mammalian cells, typically immortalized, *e.g.*, mouse, CHO, human and monkey cell lines and derivatives thereof. Preferred host cells are able to process the variant gene product to produce an appropriate mature polypeptide. Processing includes glycosylation, ubiquitination, disulfide bond formation, general post-translational modification, and the like.

The protein may be isolated by conventional means of protein biochemistry and purification to obtain a substantially pure product, *i.e.*, 80, 95 or 99% free of cell component contaminants, as described in Jacoby, *Methods in Enzymology* Volume 104, Academic Press, New York (1984); Scopes, *Protein Purification, Principles and Practice*, 2nd Edition, Springer-Verlag, New York (1987); and Deutscher (ed), *Guide to Protein Purification, Methods in Enzymology*, Vol. 182 (1990). If the protein is secreted, it can be isolated from the supernatant in which the host cell is grown. If not secreted, the protein can be isolated from a lysate of the host cells.

The invention further provides transgenic nonhuman animals capable of expressing an exogenous variant gene and/or having one or both alleles of an endogenous variant gene inactivated. Expression of an exogenous variant gene is usually achieved by operably linking the gene to a promoter and optionally an enhancer, and microinjecting the construct into a zygote. *See* Hogan et al., "Manipulating the Mouse Embryo, A Laboratory Manual," Cold Spring Harbor Laboratory. Inactivation of endogenous variant genes can be achieved by forming a transgene in which a cloned variant gene is inactivated by insertion of a positive selection marker. *See* Capecchi, *Science* 244, 1288-1292 (1989). The transgene is then

introduced into an embryonic stem cell, where it undergoes homologous recombination with an endogenous variant gene. Mice and other rodents are preferred animals. Such animals provide useful drug screening systems.

In addition to substantially full-length polypeptides expressed by variant genes, the present invention includes biologically active fragments of the polypeptides, or analogs thereof, including organic molecules which simulate the interactions of the peptides. Biologically active fragments include any portion of the full-length polypeptide which confers a biological function on the variant gene product, including ligand binding, and antibody binding. Ligand binding includes binding by nucleic acids, proteins or polypeptides, small biologically active molecules, or large cellular structures.

Polyclonal and/or monoclonal antibodies that specifically bind to variant gene products but not to corresponding prototypical gene products are also provided. Antibodies can be made by injecting mice or other animals with the variant gene product or synthetic peptide fragments thereof. Monoclonal antibodies are screened as are described, for example, in Harlow & Lane, *Antibodies, A Laboratory Manual*, Cold Spring Harbor Press, New York (1988); Goding, *Monoclonal antibodies, Principles and Practice* (2d ed.) Academic Press, New York (1986). Monoclonal antibodies are tested for specific immunoreactivity with a variant gene product and lack of immunoreactivity to the corresponding prototypical gene product. These antibodies are useful in diagnostic assays for detection of the variant form, or as an active ingredient in a pharmaceutical composition.

V. Kits

The invention further provides kits comprising at least one allele-specific oligonucleotide as described above. Often, the kits contain one or more pairs of allele-specific oligonucleotides hybridizing to different forms of a polymorphism. In some kits, the allele-specific oligonucleotides are provided immobilized to a substrate. For example, the same substrate can comprise allele-specific oligonucleotide probes for detecting at least 10, 100 or all of the polymorphisms shown in Table 1. Optional additional components of the kit include, for example, restriction enzymes, reverse-

transcriptase or polymerase, the substrate nucleoside triphosphates, means used to label (for example, an avidin-enzyme conjugate and enzyme substrate and chromogen if the label is biotin), and the appropriate buffers for reverse transcription, PCR, or hybridization reactions. Usually, the kit also contains instructions for carrying out the methods.

EXAMPLES

The polymorphisms shown in Table 1 were identified by resequencing of target sequences from eight unrelated individuals of diverse ethnic and geographic backgrounds by hybridization to probes immobilized to microfabricated arrays. The strategy and principles for design and use of such arrays are generally described in WO 95/11995. The strategy provides arrays of probes for analysis of target sequences showing a high degree of sequence identity to the reference sequences of the fragments shown in Table 1, column 1. The reference sequences were sequence-tagged sites (STSs) developed in the course of the Human Genome Project (*see, e.g., Science* 270, 1945-1954 (1995); *Nature* 380, 152-154 (1996)). Most STS's ranged from 100 bp to 300 bp in size.

A typical probe array used in this analysis has two groups of four sets of probes that respectively tile both strands of a reference sequence. A first probe set comprises a plurality of probes exhibiting perfect complementarily with one of the reference sequences. Each probe in the first probe set has an interrogation position that corresponds to a nucleotide in the reference sequence. That is, the interrogation position is aligned with the corresponding nucleotide in the reference sequence, when the probe and reference sequence are aligned to maximize complementarily between the two. For each probe in the first set, there are three corresponding probes from three additional probe sets. Thus, there are four probes corresponding to each nucleotide in the reference sequence. The probes from the three additional probe sets are identical to the corresponding probe from the first probe set except at the interrogation position, which occurs in the same position in each of the four corresponding probes from the four probe sets, and is occupied by a different nucleotide in the four probe sets. In the present analysis, probes were 25 nucleotides

long. Arrays tiled for multiple different references sequences were included on the same substrate.

Multiple target sequences from an individual were amplified from human genomic DNA using primers for the fragments indicated in the listed Web sites. The amplified target sequences were fluorescently labelled during or after PCR. The labelled target sequences were hybridized with a substrate bearing immobilized arrays of probes. The amount of label bound to probes was measured. Analysis of the pattern of label revealed the nature and position of differences between the target and reference sequence. For example, comparison of the intensities of four corresponding probes reveals the identity of a corresponding nucleotide in the target sequences aligned with the interrogation position of the probes. The corresponding nucleotide is the complement of the nucleotide occupying the interrogation position of the probe showing the highest intensity (see WO 95/11995). The existence of a polymorphism is also manifested by differences in normalized hybridization intensities of probes flanking the polymorphism when the probes hybridized to corresponding targets from different individuals. For example, relative loss of hybridization intensity in a "footprint" of probes flanking a polymorphism signals a difference between the target and reference (i.e., a polymorphism) (see EP 717,113, incorporated by reference in its entirety for all purposes). Additionally, hybridization intensities for corresponding targets from different individuals can be classified into groups or clusters suggested by the data, not defined *a priori*, such that isolates in a give cluster tend to be similar and isolates in different clusters tend to be dissimilar. See WO 97/29212 filed February 7, 1997 (incorporated by reference in its entirety for all purposes). Hybridizations to samples from different individuals were performed separately. Table 1 summarizes the data obtained for target sequences in comparison with a reference sequence for the eight individuals tested.

From the foregoing, it is apparent that the invention includes a number of general uses that can be expressed concisely as follows. The invention provides for the use of any of the nucleic acid segments described above in the diagnosis or monitoring of diseases, such as cancer, inflammation, heart disease, diseases of the CNS, and susceptibility to infection by microorganisms. The invention further

provides for the use of any of the nucleic acid segments in the manufacture of a medicament for the treatment or prophylaxis of such diseases. The invention further provides for the use of any of the DNA segments as a pharmaceutical.

All publications and patent applications cited above are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication or patent application were specifically and individually indicated to be so incorporated by reference. Although the present invention has been described in some detail by way of illustration and example for purposes of clarity and understanding, it will be apparent that certain changes and modifications may be practiced within the scope of the appended claims.

WHAT IS CLAIMED IS:

1. A nucleic acid segment of between 10 and 100 bases from a fragment shown in Table 1 including a polymorphic site, or the complement of the segment.

2. The nucleic acid segment of claim 1 that is DNA.

3. The nucleic acid segment of claim 1 that is RNA.

4 The segment of claim 1 that is less than 50 bases.

5. The segment of claim 1 that is less than 20 bases.

6. The segment of claim 1, wherein the fragment is 19201 and the polymorphic site is at position 179.

7. The segment of claim 1, wherein the polymorphic site is diallelic.

8. The segment of claim 1, wherein the polymorphic form occupying the polymorphic site is the reference base for the fragment listed in Table 1, column 3.

9. The segment of claim 1, wherein the polymorphic form occupying the polymorphic site is an alternative form for the fragment listed in Table 1, column 5.

10. An allele-specific oligonucleotide that hybridizes to a segment of a fragment shown in Table 1, column 8 or its complement.

11. The allele-specific oligonucleotide of claim 10 that is probe.

12. The allele-specific oligonucleotide of claim 10, wherein a central position of the probe aligns with the polymorphic site of the fragment.

13 The allele-specific oligonucleotide of claim 10 that is a primer.

14. The allele-specific oligonucleotide of claim 13, wherein the 3' end of the primer aligns with the polymorphic site of the fragment.

15. An isolated nucleic acid comprising a sequence of Table 1, column 8 or the complement thereof, wherein the polymorphic site within the sequence or complement is occupied by a base other than the reference base show in Table 1, column 3.

16. A method of analyzing a nucleic acid, comprising:

obtaining the nucleic acid from an individual; and

determining a base occupying any one of the polymorphic sites shown in Table 1.

17. The method of claim 16, wherein the determining comprises determining a set of bases occupying a set of the polymorphic sites shown in Table 1.

18. The method of claim 16, wherein the nucleic acid is obtained from a plurality of individuals, and a base occupying one of the polymorphic positions is determined in each of the individuals, and the method further comprising testing each individual for the presence of a disease phenotype, and correlating the presence of the disease phenotype with the base.